School of Education

Fluency Strategy Training and the L2 Oral Task Performance of Indonesian EFL Classroom Learners

Dony Marzuki

0000-0002-8284-8042

This thesis is presented for the Degree of

Doctor of Philosophy

of

Curtin University

September 2021

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

Signature : ……..

Date : 30 September 2021

Acknowledgments

First and foremost, praise and thanks the God, the Almighty, for His showers of blessings throughout my research work to successfully complete the research. I am also very grateful to LPDP (Indonesia Endowment Fund for Education) for providing the scholarship which has covered the entire study life.

I would like to express my sincere gratitude to my main research supervisor, Associate Professor Craig Lambert, for providing invaluable guidance throughout this research. His work ethic, patience, and motivation have deeply inspired me. He has taught me the best way to carry out the research I was interested in and present the research works clearly. It was a great privilege and honour to work and study under his guidance. I would also like to thank Dr. Martin Cooper, my co-supervisor, for all the reviews, comments, and suggestions to improve the thesis. My gratitude also goes to Professor Rod Ellis, my associate supervisor, for his valuable advice to conduct a research especially at the beginning of the study. It would be impossible for me to achieve this accomplishment without all of you.

I am incredibly grateful to my parents, whose love, prayers, and sacrifices have helped me up to this point in life. I also want to express my deep love and gratitude to my wife and my children for the continuing support, understanding, and prayers throughout my study time.

Finally, I would also like to thank Politeknik Negeri Padang, especially the English department's staff and students, who had provided me with the opportunity to pursue my education to the highest degree and conduct the research in the institution. Thank you very much for your supports.

Abstract

L1 speakers may produce speech fluently due to highly automatized access to a rich range of linguistics resources which allow them to focus their attention on what they want to say with little or no attention to how they say it. On the contrary, L2 learners typically lack automated access to the resources required to express what they want to say in a given situation. To produce a speech, L2 learners have to consider what they want to say and devote focused attention to deciding how to say it. A key question for instructed second language acquisition is whether explicit fluency strategy training can help L2 learners develop automated access to the linguistic resources they choose to use on a given task type.

This quasi-experimental study of two intact classes in a third-tier Indonesian university investigated the impacts of two instructional conditions on the fluency of learners' performance of a target task in the L2 syllabus. In the first condition (massed practice), learners rehearsed the versions of the target speaking task for the entire lesson. In the second condition (strategy training), learners were provided with explicit fluency strategy training for a portion of each lesson before rehearsing versions of the speaking task for the remainder of the lesson. Both instructional conditions allowed learners to benefit from language input and task repetition to help conceptualize and encode the content of their speech.

The participants in the study were second-year English learners from a third-tier higher degree institution in Indonesia, whose English proficiency level ranged from high-beginning to low-intermediate (CEFR range of A2-B1). These learners rarely spoke English inside or outside of classroom due to the status of English as a foreign language. Therefore, speaking was considered the most difficult skill to acquired.

Audio recordings of participant's oral presentations in pre-test and post-test were transcribed and coded for speech fluency measures, including speech rate, articulation rate, phonation/time ratio, number of filled and unfilled pauses, mean duration of pauses, mean length of run, verbatim repetition, and repairs. In addition, two raters also judged the pre-

tests and post-tests recordings based on the proficiency, pronunciation, discourse, vocabulary, grammar, and complexity of the speech. These data were analyzed in two ways. The first involved a group comparison based on statistical inferential and estimation analyses, and the second involved a phase transition analysis based on complex dynamic systems theory.

Non-parametric inferential analysis revealed that neither instructional condition significantly improved participants' speech fluency. On the contrary, the effect of both instructional conditions on oral proficiency as measured by teachers' rating of the performances of the respective groups did reach statistical significance. Both instructional conditions also had a comparable impact on learners' speech fluency and oral proficiency gains as suggested by the statistically non-significant results of both groups' comparison. In addition, the estimation analysis revealed that both instructional groups could potentially have a potentially meaningful improvement in developing learners' speech fluency and oral proficiency. The effect size on speech fluency improvement ranged from small to medium, while for oral proficiency the effect sizes ranged from medium to large.

For EFL learners in the context of the current study, effect sizes range from small to medium could be considered a potentially meaningful improvement for their speech fluency development. Therefore, small to medium effect sizes found on some speech fluency measures (speech rate, articulation rate, phonation time ratio, and mean duration of pauses) were considered as potentially important effects for participants' speech fluency development given the context in which the study was conducted. The estimation analysis also pointed to the possibility of differential and potentially complementary effects of implicit instruction and explicit strategy instruction. The analysis of effect sizes revealed that explicit instruction group performed slightly better than the implicit instruction group in terms of phonation time ratio and mean duration of pauses, whereas the implicit instruction group did slightly better on articulation rate than the explicit instruction group. This trade-off between the two types of instruction could mean that a combination of implicit and explicit instruction could be used to better improve learners' speech fluency.

The improvement in oral proficiency as measured by the teachers' ratings was statistically significant with meaningful effects. Individually the impact of explicit instruction on oral proficiency was very meaningful in general proficiency, vocabulary, and complexity. The effect sizes found on these three scales were large. Meanwhile, large effect sizes in implicit instruction were found in general proficiency, discourse, and complexity. With mostly small effect sizes, the explicit strategy training instruction had a slightly better impact than the implicit instruction on learners' general proficiency, vocabulary, and complexity. The impact of implicit instruction on learners' pronunciation, discourse, and grammar seemed to be better than explicit instruction with medium effect sizes. This indicates that judgements of quantitative measures of speech processing may not be reflective of the teachers' judgements or oral proficiency in the instructional context in which the study was conducted.

Meanwhile, the phase transition analysis revealed a degree of variability in participants' speech fluency development, indicating that each of the analyzed participants had different developmental patterns, which may have been partly due to their proficiency, educational background, and the instruction.

It was concluded that both instructional conditions could be applied with potentially complementary effects in EFL classrooms at the university in Indonesia at which the study was conducted, particularly with regard to increasing teachers' perceptions of learners' oral proficiency in the program. EFL teachers could also consider the additional recommendations made based on the current study to improve the application of each type of instruction in their classrooms.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

A human being is a social creature who needs to interact with others. This interaction is mainly conducted using language through speaking (speech), listening, writing, and reading. In its early development, a language was initially just a series of calls or gestures that evolved and advanced into complex communication or language (Pinker & Bloom, 1990). A language was maybe just a form of grunting or gesturing to deliver messages initially. It has now developed into a complex system of communication that requires diverse expertise to understand how it works. Although language is known to be studied under linguistics, its complexity demands a different understanding from theoretical and practical points of view. The theoretical point of view comes from psychological, sociological, and neurological studies. These studies then create new branches of interdisciplinary linguistics study of psycholinguistics, sociolinguistics, and neurolinguistics (Finch, 2016).

Each of these branches of linguistics studies plays a distinct role in language teaching and learning in the first language (L1), second language (L2), and (FL) foreign language context. As can be inferred from its name, sociolinguistics comes from the word society and linguistics. It is a branch of linguistics that studies the effects of society and all of its components like culture, norm, social status, age, gender, and others on language. Sociolinguistics also explores language varieties among societies, races, people, and learners. A sociolinguistic study is usually conducted to understand the relationship between language and society better. For example, spoken or written language samples are taken from a random population to be measured against socio-economic indices like education, income, age, or culture. Neurolinguistics, on the other hand, is different from linguistics because it studies language users rather than the language itself (Lebrun, 1983). Neurolinguistics is the study of how language is represented in the brain and how the brain processes language acquisition and retrieval (Menn, 2020). Menn also explains further that neurolinguistics study is intended to answer questions related to the sophistication of the human neural system concerning language acquisition and retrieval in everyday communication.

Meanwhile, psycholinguistics is a linguistics study that seeks to understand the mental process surrounding language acquisition and language use. Psycholinguistics plays an essential role in understanding the mental process involved in acquiring and producing a language or speech. Together, these three linguistics branches have shaped people's understanding of the language's production, the teaching and learning of a language, and language acquisition.

1.1 Speech Production

In psycholinguistics studies' language production is often related to speech production. Their relation is often defined as the relationship between knowledge and behavior. Language is the knowledge, while speech, which consists of meaningful sounds, is the behavior that reflects the knowledge. Levelt's model (Levelt, 1989, 1999) of the speech production process is widely used to understand how speech is produced. Initially, this model was developed to explain the psycholinguistic process experienced by first language (L1) or unilingual speakers. Hence, it can also be applied to understand the second language (L2) speech production. "Many aspects of speaking are the same for monolingual and bilingual speakers" (de Bot, 1992, p. 2), and "the basic psycholinguistic mechanisms involved in producing speech seem to be very similar" (Kormos, 2011, p. 40). Levelt's model posits three primary stages of speech production: conceptualization, formulation, and articulation. Conceptualization is the first stage of the process through which a speaker selects information to be conveyed and organizes it to create a 'preverbal plan.' Next, the preverbal plan is encoded lexically and grammatically during the formulation stage and then articulated phonetically to produce a phonetic plan during the articulation stage.

1.1.1 Conceptualization

According to Levelt (1989), a message is started with the creation of a communicative intention processed by the conceptualizer component of the human brain. The conceptualizer is also responsible for converting the communicative intention into a conceptual plan, also known as a preverbal plan. During this process, declarative knowledge is accessed. Declarative knowledge is knowledge about 'what' such as general facts,

personal history, the world, various situations, and the discourse record or what has already been talked about or said. Both in L1 and L2 contexts, this process needs attentional control that takes up memory resources. For example, in order to prompt a communicative intention, different types of information must be attended to so they can be retrieved from the long-term memory and stored into working memory to receive focal attention. Once the preverbal plan is created, it is ready to send to the formulator.

1.1.2 Formulation

During the formulation process, the preverbal plan accesses the lexicon to actuate items that can represent the different chunks of the intended message. This process involves grammatical and phonological encoding, which are lexically driven. The grammatical encoding requires the application of lexical access and syntactic procedures. Lexical access takes place when a chunk in the preverbal plan activates some lemmas in the lexicon, and the lemmas that receive the highest activation are chosen (Dell, 1986). Usually, lemmas will receive the highest activation if their semantic specifications match the preverbal plan concept. For example, if a man visits a café and wants to order a coffee from the waitress (the preverbal plan), then of all the active words in his lexicon, the words 'I,' 'want,' and 'coffee' probably will receive the highest activation because they best match the preverbal plan. Then, those lemmas' syntactic properties will be available and hence prompts the occurrence of syntactic building procedures. Next, the phonological encoding process will occur to select specific morphological and phonological forms for the selected lemmas. By this time, the preverbal plan has changed into a phonetic plan ready for articulation.

1.1.3 Articulation

Articulation is the process of converting the phonetic plan into speech sounds by using speech apparatus components like laryngeal, supra-laryngeal, and respiratory systems. The process starts when the generated phonetic plan is sent to and stored for a while in the working memory in the form of buffered information known as internal speech. The internal speech will prompt the motor commands to be unfolded and create the message or an overt speech as the final articulation result.

In addition, Levelt (1989) also theorizes that a monitoring process of speech production can occur during the conceptualization and articulation stages. Levelt believes that a speaker can access both the internal and the overt speech to check and fix any potential problem in the production. First, while the message is still in the form of internal speech, a speaker can match the preverbal plan with the message's intention and check the internal speech at the working memory before it is actually articulated. Meanwhile, if the message is already in the form of overt speech or is articulated, a speaker can monitor the meaning, sound, and forms of the production because the conceptualization system is equipped with audition and speech comprehension components. In case the overt speech turns out to have problems, the speaker can take several measures like revising the preverbal plan, changing some lexical items, generating a new message, or ignoring the problem.

Levelt assumes that these stages of speech production operate in parallel. If speech production processing in this parallel module is adequately automatic, the three modules are hypothesized to work simultaneously (Levelt, 1989). For example, a proficient speaker may have automatic processing in the formulation module to operate parallel with other modules. Meanwhile, the conceptualization and monitoring modules may still operate in serial processing since they require attentional resources. Ideally, the formulator is working on a piece of partially complete information when, at the same time, the conceptualizer continues to store the preverbal plan. In this case, the conceptualizer does not wait for the formulator and articulator to be at the same level to produce a speech. Proficient L1 speakers would be best positioned to practice this parallel processing because they have the necessary grammatical and phonological encoding capacity supported by automatized linguistic knowledge. This linguistic knowledge can be retrieved and implemented quickly because L1 speakers usually have a massive store of ready-made language chunks.

1.1.4 L2 Speech Production

While L1 speakers may produce an automatic speech production process, L2 speakers might have difficulties in almost every stage of the process. The conceptualization process might be entirely conscious and cognitively demanding for both L1 and L2 speakers. This process

usually takes up memory resources (de Bot, 1992) because it involves *macro-planning* and *micro-planning* the message generating process (Kormos, 2006), which requires conscious decision-making action (Roberts & Kirsner, 2000). This decision-making action involves a controlled process that is usually slow to execute (Roberts & Kirsner, 2000). During the formulation process, L2 speakers may require "a conscious search mechanism to retrieve the appropriate lemma matching the activated concept and to perform the ensuing syntactic and morphological encoding procedures" (Kormos, 2011, p. 52). This conscious mechanism may result in a slow formulation process and a breakdown in speech production. L2 speakers might also have problems in the articulation and monitoring process because these processes require speakers' conscious control or attention (Kormos, 2006). Since L2 speakers' attentional control is limited due to the working memory constraints (Gathercole & Baddeley, 2014), their attention to articulation and monitoring processes might result in slower speech production, frequent pauses, and hesitations (de Jong, 2016). In other words, the problems L2 speakers have during the speech production process will make their speech becomes less fluent.

In order to produce fluent speech, L2 speakers should make their speech production procedures efficient (de Jong & Perfetti, 2011). For this purpose, L2 speakers need to proceduralize their linguistic declarative knowledge and reduce the burden of the attentional resources by making the speech production process automatized (de Jong & Perfetti, 2011). Automatized encoding process might be facilitated by proceduralization (Towell, Hawkins, & Bazergui, 1996) and repeated practice of the language (Bygate, 1996; Gatbonton & Segalowitz, 2005). Also, comprehensible language input might enable L2 learners to produce fluent speech by reducing attentional demands and facilitating L2 development in a complementary fashion (Van Patten, 2002).

1.1.5 Proceduralization

The proceduralization might occur during the speech production's formulation process. Since the formulator is responsible for syntactic and grammatical encoding for the intended message, the proceduralization may result in the automaticity of the encoding process. The

automatized encoding process may lead to less attentional efforts for the formulation process. Towell et al. (1996) argued that proceduralization could be promoted by language use and interaction with L1 speakers in the target language environment. They also hypothesized that proceduralization might be indicated by speakers' ability to produce longer speaking time without pausing. To investigate this hypothesis, Towell et al. (1996) conducted a study involving twelve learners studying L2 in the target language environment. By analyzing learner's speech before and after the interaction (1-year duration), the study found that the learners' speaking improved. They could produce longer speech without pausing. The study also found that the improvement was caused by learners' success in proceduralizing syntactic knowledge.

1.1.6 Language Input and Repeated Practice

L2 input is comprehensible when it has features that are noticed and processed by learners (Doughty, 2008). This input can potentially change learners' interlanguage representations (Skehan, 1998), and it is useful for speech production (Van Patten, 2003). Comprehensible input may also influence learners' speech fluency because learners can use the necessary features to refine their production in order to achieve enhanced output. For example, learners may notice some language structures, terms, expressions, even readily available phrases or sentences from the input and use them as a learning output. In language learning, this output may be applied to task practice so it can further provide more benefits for learners, especially when repetition is provided. Repeated practice can improve fluency since the retrieval speed of words and phrases will increase along with the repetition (de Jong & Perfetti, 2011). Besides, much practice in speech production can make the speed of access and control of linguistic forms fluent (de Jong & Perfetti, 2011).

1.2 Speech Fluency

Fluency is one of the key features of oral performance (Ellis, 2009; Skehan, 2009). Together with complexity and accuracy, these three elements are often researched together to understand learners' speech performance. Fluency is also regarded as a reliable predictor of L2 proficiency (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012) in regards to its ability

to help understand the speech production process and even language acquisition (Kormos, 2006). As a component of oral performance, fluency has three aspects (Segalowitz, 2010) that determine how it is seen and understood. The first aspect is cognitive fluency, which refers to the underlying mechanisms of a speaker's speech production. The second aspect is perceived fluency, which refers to how a listener perceives and understands the cognitive fluency of a speaker. The last aspect is utterance fluency, also known as the measurable aspect of speech fluency, or the observable speech behaviours reflecting a speaker's cognitive fluency. Therefore, utterance fluency has been used widely in research investigating fluency in oral or written performance.

Skehan (2003) and Tavakoli and Skehan (2005) suggest that utterance fluency should be measured in terms of three sub-constructs: speed, breakdown, and repair. Speed fluency is the speed of delivery represented by the speech rate, which is derived from the number of syllables uttered per second/minute/total time (Ellis & Barkhuizen, 2005). Another measure of speed fluency is articulation rate, which is calculated as the number of pruned syllables per total time excluding silences and pauses (de Jong & Perfetti, 2011). In contrast, breakdown fluency reflects anything that disrupts speech flow, like pauses or silences (Skehan & Foster, 1999). Finally, repair fluency represents the monitoring process of speech production, which is usually reflected by repetitions, replacements, and reformulations (Skehan & Foster, 1999). Hence, “fluency of L2 performance reflects the efficient functioning of speech production mechanisms, including the automaticity of encoding processes, the conceptual demands of planning the content of the message, and the skillful handling of performance breakdowns” (Lambert, Kormos, & Minn, 2017, p. 170).

1.3 Fluency Strategy Training

L2 learners might also make their speech production process efficient if they apply learning strategies specifically designed to improve speech fluency. It is argued that learning strategies play an essential role in learning a second or foreign language (Oxford & Leaver, 1996), and they seem to be associated with second or foreign language proficiency (Ardasheva, Wang, Adesope, & Valentine, 2017; Chamot, Barnhardt, El-Dinary, & Robbins,

1999; Cohen, Weaver, & Li, 1996; Dörnyei, 1995; Oxford & Burry-Stock, 1995). Some researchers even argue that successful language learners apply specific learning strategies that could be utilized to benefit others who are less successful in their learning (Rubin, 1981). Hence, it might be possible to train learners to use learning strategies.

For the last four decades, learning strategies have been applied in language teaching and research through strategy training instruction to improve learners' language skills. Plonsky (2011), in his meta-analysis, found that most studies in strategy training instruction revealed that the training resulted in stronger effects on learners' speaking proficiency. As one critical feature of oral proficiency, speech fluency might also be improved through strategies training instruction (de Jong & Perfetti, 2011; Seifoori & Vahidi, 2012; Tavakoli, Campbell, & McCormack, 2016). Some studies have applied fluency strategy training to improve learners' reading, writing, and speaking fluency. Especially for speaking or oral fluency, several studies (e.g., Seifoori & Vahidi, 2012; Tavakoli et al., 2016) have focused on investigating learners' fluency improvement through fluency strategy training.

Except for Seifoori and Vahidi's (2012) study, most of the oral fluency and fluency strategy training studies are cross-sectional using inferential statistics. With this approach, a study can make inferences from the data collected to more general conditions. In other words, a conclusion can be made beyond the extent of the data itself. To understand how university students in Indonesia develop their language skills, a researcher needs to study groups of university students from one university that are representative of the broader educational context. The inferential statistics have been rooted in applied linguistic because they can provide "a picture of the 'grand sweep development' where global structure and similarities across participants can be seen" (Larsen-Freeman, 2006, p. 613). However, the main focus of this approach is the end product of a lesson, instruction, or treatment at the group level, not the actual learning process undergone by learners to get to the end product. Knowing this actual learning process on an individual learner basis would be beneficial for research because it can give insight regarding the actual change in progress in L2 development.

1.4 Complex Dynamic System Theory

A study could be conducted longitudinally and have more data points set in short intervals, so any changes in language use can be tracked (Baba & Nitta, 2014). The collected data can then be analyzed in a qualitative manner by choosing one or more learners as the focus of the analysis. This kind of study can be carried out from a Complex Dynamic System (CDS) perspective. CDS, or what is popularly known as complex dynamic system theory (CDST), is "a set of concepts that describe behavior as the emergent product of a self-organizing, multi-component system evolving over time" (Perone & Simmering, 2017, p. 4). CDST may provide "a powerful metaphorical model of applied linguistic processes, allowing holistic descriptions of situated phenomena, and addressing the connectedness and change" for language studies (Cameron, 2015). Therefore, language use and language development can be understood beyond the nature of hypotheses, causality, and prediction (Larsen-Freeman & Cameron, 2008, p. 203).

CDST views the physical world elements, including language, as complex systems regarded as emergent assemblies of interconnected components that form a connected whole through interaction (Van Geert, 2008). CDST considers language a system integrated into numerous sub-systems in complete inter-connectedness (Lowie & Verspoor, 2015). This system is constantly changing and is "hardly ever stable and cannot be characterized as predetermined and linear" (Lowie & Verspoor, 2019). Although a system usually changes in a continuous manner in a particular state, it can sometimes experience a radical or discontinuous change which causes the system to undergo a phase transition and reorganization into a higher-level state (Van Dijk & Van Geert, 2007). In a CDST approach, emergence is a general principle for explaining developmental change (Lewis, 2000). Therefore identifying emergence in learner's language use over a period of time and clarifying the ways the whole language system operates around this critical moment becomes the primary goal of research based on this approach (Baba & Nitta, 2014).

One study that employed both quantitative and qualitative approaches of CDS is the study of Larsen-Freeman (2006). The study investigated the writing and speaking performances of

five L2 learners on the same task. The performance was repeated four times over six months. The study revealed that despite the steady increase of complexity, accuracy, and fluency of the learners' productions at the group level, the individual production did not "follow the same developmental pathway" (Larsen-Freeman, 2006). There was considerable intra-and inter-individual variation. However, unlike the present study, the focus of Larsen-Freeman's (2006) study is not speech production or oral fluency.

To the best of the author's knowledge at the time of conducting the present study, the only study that investigated L2 learners' fluency by using the CDST approach was the study of (Baba & Nitta, 2014). It was a case study involving 2 Japanese learners and analyzing 30 samples of students' writing completed in one academic year. The purpose was to investigate the emergence of phase transitions, indicating a sudden jump, anomalous variance, divergence, and qualitative change in the data. The result showed that both students experienced at least one phase transition (emergence) with different characteristics at a different time.

1.5 The Present Study

Based on the theory and research summarized above, it was hypothesized that to produce fluent speech, L2 speakers should develop a degree of automaticity in the speech production process or learn and practice the fluency strategies through fluency strategy training. Both methods could be applied in classroom learning by using specific approaches to instruction. The automaticity, which could be achieved through proceduralization, repeated practice, and comprehensible input, might be learned in the L2 classroom using implicit task-based instruction. Meanwhile, fluency strategy training is usually conducted through explicit instruction.

Th present study investigated the effects of two L2 classroom instructional conditions, explicit fluency strategy training and implicit task-based instruction, on developing L2 speech fluency of low-to-intermediate-level learners of English as a foreign language at a third-tier university in Indonesia. One group of learners was trained to use strategies specifically designed to improve their speech fluency through explicit instruction. These

learners were also provided the opportunity to benefit from language input and task repetition that can help conceptualize and encode the content of their speech. Finally, the learners could employ their linguistic repertoire, including those they just received from the input, by repeatedly performing related-monologic tasks. Meanwhile, the other group was provided the same learning opportunity through implicit task-based instruction without strategy training.

The following research questions guided the study:

1. Does explicit fluency strategy training instruction impact L2 learners' speech fluency?
2. Does implicit instruction in task-based form impact L2 learners' speech fluency?
3. How does L2 speech fluency change over ten weeks of instruction?
4. Do the two forms of instruction impact L2 learners' general oral proficiency base on teachers' judgment?

To answer these questions, oral performances from 54 learners were recorded and transcribed following the procedures of Foster, Tonkyn, and Wigglesworth (2000) for the transcription of oral L2 discourse into Analysis of Speech Units (AS-unit). The speech fluency measures were based on utterance fluency, consisting of three main characteristics (Skehan, 2003; Tavakoli & Skehan, 2005): speed, breakdown, and repair fluency. Speed fluency was represented by *Speech Rate, Articulation Rate, and Mean Length of Run* (three measures). Breakdown fluency includes *Filled and Unfilled Pauses* within and between clauses (four measures), and Repair was determined by *Repetitions, False Starts, Reformulations, and Replacements* (four measures). Another measure that dealt with the quantity of the speech, *Phonation Time Ratio,* was also investigated. These are considered to be some of the most reliable measures of utterance fluency (Witton-Davies, 2014).

With the exception of Research Question 3, statistical inferential and estimation analyses with effect sizes and confidence intervals were employed to answer the research questions. Non-parametric tests, combined with estimation analysis and teachers' ratings were used to analyse the performance data as the scores on the dependent variables for each group were not normally distributed. This was due to the limited number of participants, extreme

values, and a large number of values close to zero or a natural limit. The pre-test and post-test results of the utterance fluency measures were analyzed and the results were compared between the two instructional groups.

Meanwhile, to answer Research Question 3, two participants with high fluency gains from each group of the study were selected and analyzed individually using phase transition analysis of the CDS approach. The four participants' data were analyzed to find any sign of phase transitions in the development of their speech fluency. The phase transition analysis included three criteria: (1) a sudden jump or the abrupt change in the system from one state to another, (2) anomalous variance or a phenomenon that happen in the system when it becomes unstable and shows an increasing variability as the transition approaches, and (3) and qualitative change or a system reorganization which qualitatively differentiate the current mode from the previous one (Baba & Nitta, 2014). The sudden jumps were identified using change-point analysis, indicating the points at which a sudden change occurs in the data (Steenbeek, Jansen, & van Geert, 2012). The change point analysis used calculation derived from Change-Point Analyser software (Taylor, 2000). The anomalous variance was analyzed by using a line graph of the moving standard deviation of residuals to visualize the greatness of fluctuations in text length (Van Geert & Van Dijk, 2002). The transcription of participants' speech performances was assessed qualitatively to find the indications for the qualitative change.

To date, no studies to the author's knowledge investigating speech fluency have applied this combination of methods. The estimation approach was used to gather data explaining the effects of treatments at the group level, which could be generalized to a broader population with a similar context. The case study under the CDST approach could shed light on individual developmental changes over time (Thelen & Smith, 1994). Therefore, the present study was intended to improve L2 teachers' and researchers' understanding regarding the effects of both instructional conditions on L2 learners' speech fluency development on a general and individual basis.

1.6 Structure and Organization of the Thesis

Following this introductory chapter are eight chapters that detail the research. Chapters 2 and 3 discuss current perspectives regarding the role of speech fluency in theories and practices of L2 instruction. Chapter 2 details speech fluency and its' connection to the L2 speech production process. This chapter also summarizes previous studies on L2 speech fluency. Chapter 3 discusses the roles of strategies training in developing L2 speech fluency, including the application of the training in EFL classrooms. Chapter 4 then reviews and operationalizes the speech fluency measures as the dependent variables in the primary study. These are measures of speech fluency that have been used in studies of L2 oral performance.

Chapter 5 discusses the complex dynamic systems theory as the basis for the case studies as the secondary analysis in this study. The discussion includes the basic characteristics of a complex system and the phase transition analysis to analyze meaningful changes in a system's development. Chapter 6 presents the research methods that specify the research questions and outlines the study's particulars in sufficient detail to be replicable for future research.

Meanwhile, Chapters 7 and 8 describe the results of the study. Chapter 7 details the inferential and estimation analyses results, including the effect sizes and the confidence of intervals. Chapter 8 then details the phase transition analysis results that explored learners' development on an individual basis. Chapter 9 is the discussion of all results of the study as detailed in Chapter 7 and Chapter 8. First, the overall findings for the study following each research question are highlighted. Then, the implications of these findings for theories and practice in EFL instruction are described. Chapter 9 also details the limitations and recommendation of the study for future research. This chapter is closed with a conclusion for the study.

Chapter 2: Speech Fluency

This chapter discusses speech fluency as the central aspect investigated in the study. The discussion is divided into two sections. Section 2.1 highlights the fluency and the many definitions attributed to it. This section also tries to connect speech fluency and the L2 speech production process proposed by Levelt (1989, 1999), including how speech production could lead to automatic speech processing. Section 2.2 discusses researchers' and language practitioners' attempts to improve L2 speech fluency through research and practice.

2.1 Defining Speech Fluency

One noticeable difference between the first language (L1) and second language (L2) speakers is the speed with which they talk (Kormos, 2006). L2 speakers usually produce speech slower than L1 speakers' because L2 speakers need to pay attention to the entire speech production process and the lack of linguistic repertoire such as lexis, syntax, morphology, and phonology (Kormos, 2006). Relating to Levelt's (1989, 1999) speech production modules, both L1 and L2 speakers requires attention to performing the first stage of the production process, conceptualization, and monitoring. However, while L1 speakers may automatically perform the formulation and articulation stages (de Bot, 2003), L2 speakers tend to do it consciously. Automaticity has enabled L1 speakers to perform parallel processing in their speech production, which results in rapid speech performance, also known as speech fluency.

Fluency in speech production is an essential factor of first (L1) and second (L2) language acquisition. Together with accuracy and complexity, fluency has been determined as a reliable predictor of L2 proficiency (de Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013). Fluency is also considered a useful tool to comprehend the speech production process (Kormos, 2006; Segalowitz, 2010). Due to its essential role in L2 teaching and learning, fluency has gained the attention of many L2 researchers (Tavakoli et al., 2016) to understand what fluency is (Kormos, 2006; Lennon, 2000; Segalowitz, 2010; Skehan, 2009). Previous studies have tried to find ways to make L2 learners able to perform fluently in their

speech performance (Freed, 2000; Gatbonton & Segalowitz, 1988; Kormos & Dénes, 2004; Lambert et al., 2017; Nation, 1989; Seifoori & Vahidi, 2012; Tavakoli, 2010; Tavakoli et al., 2016; Tavakoli & Skehan, 2005).

Fluency can be defined based on the scope of its application. Lennon (1990, 2000) defines fluency in broad and narrow senses. In a broad sense, fluency is regarded as the reflection of general proficiency, which includes the accuracy and complexity of production (de Jong & Perfetti, 2011). In a narrow sense, fluency is just one component of proficiency, usually used as one score for assessing language skills (Kormos, 2006). Concerning the narrow sense, Lennon (1990) argues that speech fluency is the listener's impression regarding the speaker's easiness and efficiency in producing speech. By defining speech fluency based on this speaker-listener relationship, Rehbein (1987) also proposes a similar argument. According to him, speech fluency is determined by the listener's expectation of the speaker's speech production. Also, Schmidt (1992) offers a more specific definition of speech fluency in this narrow sense. He argues that speech fluency is an automatic procedural skill. All of these narrow senses of speech fluency share similarities with the definition provided by Koponen and Riggenbach (2000), Segalowitz (2010), and Skehan (2014b) that stress automaticity in speech production.

Speech fluency is then defined as flow, continuity, automaticity, or smoothness of speech (Koponen & Riggenbach, 2000). Segalowitz (2010) and Skehan (2014) also offer similar definitions, which refer to speech fluency as automaticity in learners' speech production, which is reflected in the smoothness of the speech (Skehan, 2014b). Hence, to provide a basis for the study, a more technical definition that covers automaticity, as offered by Lambert et al. (2017), will be used. Speech fluency is defined as “the efficient functioning of speech production mechanisms, including the automaticity of encoding processes, the conceptual demands of planning the content of the message, and the skillful handling of performance breakdowns” (p. 170).

In addition to the above definitions, Segalowitz (2010) offers a more technical description of speech fluency and conceptualizes it into three distinct, interrelated concepts: cognitive

fluency, perceived fluency, and utterance fluency. *Cognitive fluency* is understood as the cognitive mechanism underlying the speaker's speech performance or production. *Perceived fluency* refers to how a listener sees and understands the cognitive fluency of a speaker. *Utterance fluency* reflects a speaker's cognitive fluency, also known as the measurable aspects of speech fluency, including speed, pauses, and repairs (Tavakoli et al., 2016). utterance fluency has been used widely in research investigating fluency in oral or written performance (see Chapter 4 for the discussion).

2.1.1 Speech Production and Speech Fluency

The speech production process proposed by Levelt (1989, 1999) might provide the basis to understand speech fluency. The model posits three main information-processing stages: *conceptualization, formulation,* and *articulation*. Conceptualization is the first stage of the process through which a speaker selects information to be conveyed and organizes it to create a 'preverbal plan.' This preverbal plan will then undergo a formulation stage that includes the syntactic and morphologically encoding process to produce 'internal speech' or a 'phonetic plan.' In the articulation stage, the phonetic plan is converted into overt speech (de Bot, 1992).

Levelt (1989, 1999) theorizes that the three stages or modules of speech production operate in parallel. If speech production processing in parallel modules is adequately automatic, the three modules are hypothesized to work simultaneously. In parallel processing, the conceptualization process occurs continually by providing the preverbal plan to the formulator to begin the encoding process. When it is ready, the information is sent to the articulator to be articulated. At the same time, the conceptualizer keeps working on creating a new preverbal plan, and it does not wait for the formulator and articulator to be at the same level as it is in the entire process. For instance, to produce three sentences, the formulator begins to process the first sentence immediately and send the result to the formulator. While the formulator encodes the first sentence, the conceptualizer starts the process of the second sentence. In the meantime, the formulator already finishes with the first sentence to be sent to the articulator. While the first sentence is articulated, the

conceptualizer already works with the third sentence as the formulator might have finished with the second sentence. As for the monitoring process, Levelt (1989, 1999) theorizes that this process occurs three times at two different stages. The first is at the conceptualization stage, when the preverbal message is ready to send to the formulator. The second and the third ones occur at the articulation stage, before and after message articulation. This whole process continues until the entire speech is articulated. When speech production follows this automatic parallel processing, it will produce fluent speech.

First, to relate Levelt's speech production model with speech fluency, a connection between the model and L2 acquisition needs to be established (Towell et al., 1996). The Adaptive Control of Thought-Revised (ACT-R) (Anderson, 2000) model of cognitive development is useful for explaining the connection. As summarised by Towell et al. (1996), this model argues that learning occurs as the process of transforming factual (declarative) knowledge into operational (procedural) knowledge. This process is known as proceduralization. This model posits that knowledge is stored in three memory storages: two long-term memories, and one working memory with limited capacity. The working memory functions as a gate that connect the long-term memories with the outside world (Anderson, 1983). All information entering and leaving long-term memory must be first processed by working memory. However, working memory is a limited-capacity memory store. Therefore it cannot process a large amount of information at one time. Towell et al. (1996) argue that knowledge should be processed in procedural form instead of declarative form so working memory will not be over-loaded. However, L2 speakers, especially those who are less proficient, "may still rely on declarative knowledge they need to proceduralize" (de Jong & Perfetti, 2011, p. 537).

2.1.1.1 Declarative Versus Procedural Knowledge.

Anderson (1983) argues that all knowledge acquired by a human being, including language acquisition, is initially declarative in nature and known as declarative knowledge, or knowledge of "that" (de Jong & Perfetti, 2011). Later, through learning or practice, this knowledge becomes fully internalized as procedural knowledge (Crookes, 1991) or

knowledge of “how” (de Jong & Perfetti, 2011). Declarative knowledge includes knowledge of word forms and explicit grammatical or phonological rules. This knowledge is associated with lexical and syntactical retrieval (Kormos, 2006). Usually, assessing declarative knowledge requires speakers’ attention and takes up much more working memory space, with limited capacity. Procedural knowledge is also known as production rules. Production rules consist of lexically and grammatically encoded morphemes. Procedural knowledge can be accessed fast and automatically according to the production goals (Kormos, 2006), and a complete production can be accessed at the same time (Towell et al., 1996). This knowledge does not require much of the speaker’s attention or a large working memory capacity (Kormos, 2006). In speech production that requires rapid performance, converting this knowledge into procedural knowledge is essential (Levelt, 1989).

The difference between declarative and procedural knowledge can be understood based on L2 speakers in the EFL context. These speakers may already know that the adverbs “too” and “so” have similar meanings but different functions. The former is used for negative expression while the latter is positive. With this information in their mind, these speakers already have declarative knowledge of adverbs. With practice, this knowledge will transform into a production rule that can be used to produce speech. For example, when they need to express something extraordinary, they will use the adverb “so.” In the end, this rule will be applied automatically without requiring conscious attention (Kormos, 2006).

The conversion of declarative knowledge into procedural knowledge undergoes three stages: the cognitive stage, the associative stage, and the autonomous stage (Anderson, 1983). In the cognitive stage, all knowledge is declarative, which imposes slow interpretive mechanisms to process. Then, in the associative stage, the knowledge becomes partly declarative and partly procedural. Therefore, the access is slightly faster. In the autonomous stage, the knowledge becomes fully procedural and can be accessed very rapidly. This last stage involves the proceduralization process through “transforming the declarative knowledge into procedural knowledge to create a production which is activated as a whole” (Towell et al., 1996, p. 89). Proceduralization involves creating production rules and combining smaller production rules into larger ones” (de Jong & Perfetti, 2011). Production

rules enable speech production to be executed fast and automatically. Thus, it can be argued that proceduralization is one of the conditions that allow the occurrence of an automatic process (automaticity) in speech production (Kormos, 2006). Anderson (1983, 1995) argues that other than the faster application of rules, the development of automaticity also involves the withdrawal of attention from rule-based processing normally happens during the encoding process in the formulator. Fast and autonomous speech production as the result of proceduralization and automaticity in the encoding process is an indication of speech fluency (Towell et al., 1996).

2.1.1.2 Proceduralization and Automaticity in Speech Production Process.

For both L1 and L2 speakers, the conceptualization process might be entirely conscious and cognitively demanding. This process usually takes up memory resources (de Bot, 1992) because it involves *macro-planning* and *micro-planning* the message-generating process (Kormos, 2006). This message-generating process requires conscious decision-making action (Roberts & Kirsner, 2000). Kormos (2006) explains that the macro-planning process includes elaborating communicative intention expressed by *speech acts* such as requesting, directing, apologizing, and informing. A speaker needs to decide this speech acts before the micro-planning can be started. In micro-planning, the speaker needs to make other decisions regarding the perspective and the accessibility status of the message. A speaker could make a decision about whether he or she should say “your car is faster than mine” or “my car is slower than yours” for the perspective of the message. The accessibility status involves the speaker’s consideration of whether to choose a noun, or pronoun (e.g., “the boy” or “he”) for the message. This accessibility status is determined by the speaker’s familiarity with the words. Microplanning also involves the process of giving propositional content, and language-specific information for the message (Kormos, 2006). When all of these decisions have been applied to the intended message, the message then changes form to a preverbal plan, containing all the necessary information to convert meaning into the language (Kormos, 2006). The conceptual planning of speech (macro-planning and micro-planning) requires attention because decision-making processes involve a controlled process that is usually slow to execute (Roberts & Kirsner, 2000).

During the formulation process, L1 speakers can automatically perform the encoding processes; therefore, L1 speech production is faster than L2 speech production (Kormos, 2006). For L1 speakers, all the rules applied in the encoding processes have been proceduralized (Towell et al., 1996) due to their extensive language use. On the other hand, L2 speakers may require "a conscious search mechanism to retrieve the appropriate lemma matching the activated concept to perform the ensuing syntactic and morphological encoding procedures" (Kormos, 2011, p. 52). This conscious mechanism may result in a slow formulation process and a breakdown in speech production. This breakdown is characterized by grammatical and lexical errors, repetition of the same words with or without modification, and pauses, either filled or unfilled, within and between clauses (de Jong & Perfetti, 2011). Therefore, L2 speakers need to proceduralize their explicit phonological and grammatical rules to become procedural. Once this happens, L2 speakers can automatically perform the encoding process quickly and in parallel with other processes because procedural knowledge puts less of a burden on the working memory, which is limited in capacity (de Jong & Perfetti, 2011).

L2 speakers, especially those at the earlier stage of language acquisition, might also have problems in the articulation and monitoring process (Kormos, 2006). L2 speakers rely on their declarative knowledge to monitor their speech production (Levelt, 1989). The monitoring process also demands their attentional resources (Kormos, 1999). Meanwhile, the articulation process, which is not supported by proceduralization, might also require speakers' conscious control or attention (Kormos, 2006). Since L2 speakers' attentional control is limited because of the working memory's constraints (Gathercole & Baddeley, 2014), their attention to articulation and monitoring process might result in slower speech production, frequent pauses, and hesitations (de Jong, 2016).

Therefore, in order to achieve speech fluency, L2 speakers should make speech production procedures efficient (de Jong & Perfetti, 2011). For this purpose, L2 speakers need to proceduralize their linguistic declarative knowledge and reduce the burden of the attentional resources by making the production process automatized (de Jong & Perfetti, 2011). Hence, speech fluency is defined as the efficient functioning of speech production

mechanisms, including the automaticity of encoding processes, the conceptual demands of planning the content of the message, and the skillful handling of performance breakdowns (Lambert et al., 2017).

2.2 Improving L2 Speakers' Fluency

Most of the research on L2 fluency is focused on finding solutions to make speech production fluent. Fluent production is indicated by two interrelated conditions: automatized encoding process and less attentional effort (Kormos, 2006). It is argued that the automatized encoding process, which could also reduce attentional demand, might be facilitated by proceduralization (Towell et al., 1996). Repeated practice of the language might also help make the encoding process automatic (Gatbonton & Segalowitz, 2005) and reduce the attentional effort (Bygate, 1996). Also, comprehensible language input might enable L2 learners to produce fluent speech by reducing attentional demands and facilitating L2 development in a complementary fashion (Van Patten, 2002).

Towell et al. (1996) argued that proceduralization might occur during the formulation process of speech production. Since the formulator is responsible for syntactic and grammatical encoding for the intended message, proceduralization may result in the automaticity of the encoding process. The automatized encoding process may lead to less attentional effort during the formulation process. Towell et al. (1996) argued that proceduralization could be promoted by language use and interaction with L1 speakers in the target language environment. They also hypothesized that proceduralization might be indicated by speakers' ability to produce longer speaking time without pausing. To test this hypothesis, Towell et al. (1996) conducted a study involving twelve learners studying L2 in the target language environment. By analyzing learner's speech before and after the interaction (1-year duration), the study found that the learners' speaking improved. They could produce longer speech without pausing. The study also found that the improvement was related to learners' success in proceduralizing syntactic knowledge.

In addition, providing learners with an opportunity to repeat their oral performance might also improve their speech fluency (Ahmadian & Tavakoli, 2011; de Jong & Perfetti, 2011;

Ellis, 2005a). It is possible for L2 learners to use previously conceptualized task content and activate recently used linguistic constructions to express their message through task repetition (Lambert et al., 2017). Task repetition might “reduce the speakers’ attentional demands to conceptualize, encode, and monitor the messages simultaneously” (Lambert et al., 2017, p. 171). de Jong and Perfetti (2011) investigated the effects of task repetition on two groups of learners within a classroom. One group was assigned to do one task three times, while the other had to do three tasks with one performance for each. The results showed that task repetition had a noticeable impact on gains in different measures of fluency in the group who performed the same task three-time, whereas the other group only had the gains for a short time. de Jong and Perfetti (2011) also related their findings with proceduralization, and they argued that proceduralization denoted a modification in principle cognitive mechanisms, which lead to improvement in oral fluency. Task repetition may then be connected with the occurrence of the proceduralization process.

Also dealing with task repetition, Lambert et al. (2017) investigated the effect of the repetition of aural-oral monologue tasks on immediate gains in L2 fluency of Japanese university students in a classroom setting. The students were assigned to complete three oral tasks six times. Results revealed that task repetition was related to gains in oral fluency regardless of proficiency level or task type. Meanwhile, Bui, Ahmadian, and Hunter (2019) used five task repetition conditions in their study and investigated the effects on EFL learners' oral performance (fluency, accuracy, and complexity). The participants were divided according to the five different task repetition conditions: immediate, one-day, three-day, one-week, and two-week time-space repetitions. The study found that all conditions of task repetition had a positive effect on oral performance, with speed fluency benefitting the most from immediate and small intervals of time-space between initial and repeated performance.

In relating to the role of comprehensible input in promoting speech fluency, Skehan (1998) argues that language input can potentially change learners’ interlanguage representations. This change may be applied in their speech production process (Van Patten, 2003) to produce fluent speech. Comprehensible input may also influence learners’ speech fluency

because they can use the necessary features noticed and acquired from the input to refine their production to achieve enhanced speaking output. A study conducted by Munoz (2014) investigated the role of input and starting age in foreign language learning. The study compared the influence of different kinds of input measures such as number of years of instruction, number of hours of curricular and extracurricular lessons, and the number of hours spent abroad in an English-speaking country. One hundred sixty learners were involved and asked to retell a movie. The performances were recorded and analyzed in terms of fluency, lexical diversity, and syntactic complexity. The result showed that input had a stronger association with oral performance measures than starting age and cumulative exposure. High-quality input was found to be a good predictor of learners' oral performance. For fluency, the study found that the number of hours spent abroad was a major predictor of learners' oral fluency improvement.

For example, Zhang's (2009) case study of Chinese children investigated the role of input, together with interaction and output, in the development of oral fluency in the EFL context in China. Fifteen students were asked to conduct an oral fluency test regarding their past activity. The result was analyzed using the RSA Test battery, including the range, accuracy, appropriateness, length, and complexity of learners' production. The study focused on two participants whose oral fluency test scores were the highest. The two participants were then asked to perform the second oral fluency test to retell information from a reading passage. After comparing with the first test, the results showed that the two participants improved in terms of speech fluency in the second test. Zhang (2009) concluded that non-native oral fluency could be obtained through efficient and effective input, interaction, and output in the EFL context.

2.3 Summary

This chapter has explained speech fluency and its connection to the speech production process. It is believed that the speech production process could be made automatized through proceduralization, repeated practice, and comprehensible language learning input. Automatized speech production is one requirement to achieve speech fluency. Some studies

have suggested that L2 speech fluency can be improved through the proceduralization of language knowledge, task repetition, and comprehensible input within or outside the classroom boundary. However, a longer-duration (longitudinal) study by providing more opportunity for learners to use the language would be beneficial to yield a better result of automatization (de Keyser, 2007).

Chapter 3: Fluency Strategy Training

This chapter discusses the roles of learning strategy and training in developing speech fluency. The discussion starts with Section 3.1, highlighting learning strategies in general. Section 3.2 then discusses the application of learning strategies for language learning. This section also connects language learning strategies and fluency strategies training inside the classroom setting. Some examples of research in language strategy training are presented in section 3.3.

3.1 Learning Strategies

Driving a car or riding a motorcycle are examples of activities conducted daily. Like other daily activities, driving or riding might already be an autonomous skill for many people since they could do it very quickly and smoothly. However, when traveling to the same destination from an almost similar departure place, people tend to come up with a different completion time. Some people might spend 30 minutes, while others could be 25 or 35 minutes. Usually, these differences might be caused by the different approaches they applied during the journey. Some of them could have taken different routes or applied different speeds. The action of taking different routes or applying specific speeds in traveling may be considered as strategies in driving or riding. In this case, the strategies applied could result in people reaching their destinations faster. In other words, strategies enable them to achieve a better result.

People use strategies to make their activities more efficient and effective. Strategy application is a conscious process because the people who apply it need to analyze the activity carefully, including measuring the potential problem that will be faced with and the solution for the problem. Once this problem and solution are understood, the strategies can be applied, and more practice could improve the strategy application. Improved strategies lead to improved results. Next, the strategies could also be passed to other people so they could also improve the result of their activity.

Strategies could be applied to many kinds of activities, including learning. Learning requires learners to complete learning objectives or learning goals. Usually, learners need to apply

different strategies to achieve learning goals. Different learners tend to use different strategies in learning, and different strategies tend to give different results (Chamot et al., 1999). Some strategies might be helpful for learners, while others might depend on the learners' characteristics, tasks, and learning contexts (Chamot et al., 1999). It is argued that learning strategies might help learners perform better in achieving learning objectives (Brown, 2006; Oxford & Burry-Stock, 1995).

3.2 Language Learning Strategies

In the area of language learning, especially in the second (L2) or foreign (EFL) language contexts, learning strategies have been the object of a large body of research for the last three decades (Plonsky, 2011). The research has been directed to examine language strategies used by L2 learners during the language learning process to understand learners' metacognitive, cognitive, social, and affective processes involved in language learning (Chamot, 2005). Grenfell and Harris (2002) argue that when less successful language learners learn about learning strategies and apply them in their own learning, they can become better language learners.

Researchers in L2 learning have tried to define and operationalize language learning strategies to provide a precise definition. The initial attempts were focused on investigating what constitutes a good language learner (Naiman, Frohlich, Stern, & Todesco, 1978) and identify strategies they use to benefit their L2 learning (Rubin, 1979). An early definition of language learning strategies was provided by Rigeney (1978), who defines language learning strategies as learners' conscious steps or behaviors to enhance the acquisition, storage, retention, recall, and use of new information in learning. Similarly, Oxford and Crookall (1989) also refer learning strategies as learners' use of steps, behaviors, actions, or techniques to enhance and facilitate language acquisition. More generally, Brown (2006) defines learning strategies as processes that may contribute directly to learning, while Chamot (2004, p. 14) defines learning strategies as "the conscious thoughts and actions that learners take in order to achieve a learning goal." This definition contains some principles

involving learners' metacognitive knowledge regarding their own thinking and learning approaches (Chamot, 2004, p. 14).

3.2.1 Metacognitive knowledge

In general, metacognitive knowledge is knowledge about learning (Wenden, 1998). Flavell and Wellman (1977) originally defined metacognitive knowledge as the relatively stable information humans have about their own cognitive processes. Knowing about general strategies of doing different tasks, the condition these strategies might be used, the extent to which the strategies are effective, and knowing about self are parts of metacognitive knowledge (Flavell, 1979; Pintrich, Wolters, & Baxter, 2000). For example, learners may possess strategies for doing an oral presentation and strategies to monitor and check their performance as they speak. Furthermore, Pintrich (2002) argued that in doing a task, learners with metacognitive knowledge might actuate relevant knowledge regarding their strengths and weaknesses in relating to the task. These learners would also activate their motivation for completing the task. If learners are familiar with the topic of the task, they may consider this as their strength. If this familiarity makes them interested in doing the task, their motivation may also be enhanced. When they realize that they have the strength and the motivation to do the task, they can change their approach to the task, like devoting more time for rehearsal or planning the content for the presentation. By having more opportunities for rehearsal, learners might be able to practice different approaches to perform the presentation and choose the approach that they consider useful for the task. Hence, metacognitive knowledge could affect how learners prepare for different kinds of oral tasks.

Metacognitive knowledge is vital for developing language learning strategies because it holds the key to make learners aware of the particular cognitive strategies. Learners would be able to move toward any learning actions if they possess metacognitive knowledge. Learners with metacognitive knowledge would be able to select learning strategies for specific learning objectives and apply those strategies to different learning conditions.

Metacognitive knowledge is also recognized as “the engine that drives self-directed learning” (Knowles, 1975). Self-directed learning is the process in which learners can take initiatives for their learning, including identifying the needs, formulating the goals, identifying resources, selecting and practicing strategies, and evaluating the outcomes of the learning. Hence, it may be argued that metacognitive knowledge is necessary for language learning strategies because it activates learners' awareness of their own learning and makes them self-directed learners (Oxford, 2016).

3.2.2 Types of Language Learning Strategies

Language learning strategies have been categorized into three types: cognitive, metacognitive, and socio-affective strategies (Ardasheva et al., 2017; O'Malley, Chamot, Stewner-Manzares, Russo, & Küpper, 1985). Cognitive strategies deal with tacit behaviors and mental processes during learning (Cohen & Wang, 2018); O'Malley et al. (1985); (Oxford, 2016). These strategies are limited to specific learning tasks and involve direct manipulation of the learning material (Brown, 2006). Some examples of these strategies related to developing oral proficiency are rehearsal, note-taking, imagery, keyword, and transfer (Ardasheva et al., 2017; O'Malley et al., 1985). Rehearsal is practicing an overt speech or oral presentation before the actual performance. Note-taking involves writing down the main ideas, critical points, outline, or summary of information to be presented orally (Wenden & Rubin, 1987). Imagery is the process of relating new information to visual concepts in memory via familiar and easily retrievable visualization (Kojima, 2001). The keyword is the strategy conducted by remembering a new word in L2 by connecting it to new words to familiar words. Meanwhile, transfer involves using one’s own linguistic knowledge to complete new language learning tasks.

‘Metacognitive strategy’ is a term used to indicate an ‘executive’ function strategies (Purpura, 1997) that deal with conscious awareness during learning (Ardasheva et al., 2017; Cohen & Wang, 2018; O'Malley et al., 1985). These strategies refer to the thought and actions that can help learners think about their learning process, make a plan for learning, monitor how learning is taking place, and appraise the learning outcome (Cohen & Wang,

2018; O'Malley & Chamot, 1990). Some of the most utilized examples of metacognitive strategies are functional planning, self-monitoring, and evaluation. O'Malley et al. (1985) explain that planning covers the activity of planning for and rehearsing linguistic components necessary in performing a language task. Self-monitoring covers activities of correcting the task performance in terms of language features such as grammar, pronunciation, vocabulary, or the performance's appropriateness. Meanwhile, evaluation deals with checking the outcomes of task performances or language learning (O'Malley et al., 1985).

In contrast, socio-affective strategies have to do with social-mediating activity and interacting with others (Brown, 2006). Asking for help, self-encouragement (Hassan et al., 2005), help-seeking, time-gaining, self-solving (Dörnyei & Scott, 1997), cooperation, and asking for clarification (O'Malley et al., 1985) are some examples of socio-affective strategies. These strategies are also known as communicative strategies (Brown, 2006), usually practiced by L2 learners in various types of communication.

3.2.3 Language Learning Strategy Training

Researchers of language learning strategies agree that less successful L2 learners might be taught some of the learning strategies to help them become more effective language learners (Chamot, 2004). In its implementation, learners can learn how to use the strategies in the form of strategy training instruction (Cohen, 2003). This strategy training often involves an awareness-raising instructional model targeting task-specific strategy clusters across metacognitive, cognitive, and socio-affective strategies (Ardasheva et al., 2017). The awareness-raising model usually covers four steps: consciousness-raising, modeling, guided practice, evaluation, and goal setting (Ardasheva et al., 2017). Consciousness-raising is intended to make learners reflect on their learning and their current and potential strategies (Grenfell & Harris, 2002). In modelling, learners are presented with the new strategies and the model to apply the strategies to achieve the learning goal (Grenfell & Harris, 2002; O'Malley & Chamot, 1990). Then, learners are provided with opportunities to practice strategies in guided practice (Grenfell & Harris, 2002; O'Malley & Chamot, 1990). Finally, in

evaluation and goal setting, learners are encouraged to identify problem areas, select strategies that might help solve the problem and evaluate their learning success (Graham & Macaro, 2008). Ellis and Shintani (2014) conclude that learners' training activities in L2 classrooms often involves a combination of strategy training and awareness-raising. Cohen (2003) adds to that by mentioning that the combination of the two has become the preferred approach to learners' strategies training.

3.3 Research on Language Learning Strategies

Some research has been conducted in an attempt to relate the above strategies to improvement or development in L2 listening, reading, writing, and speaking. Initially, strategy training research was devoted to identifying the types of strategies learners use and the ways they use those strategies in conjunction with other variables such as gender, proficiency, motivation, and beliefs (Plonsky, 2011). Later, researchers tried to find ways of transferring the strategies to L2 learners in their language learning through strategy instruction or training. Usually, this research was conducted in experimental studies and has been recognized to have significant theoretical and practical potential in L2 teaching and learning ever since (Ardasheva et al., 2017; Ellis, 2005b; Grenfell, 2007; Plonsky, 2011). However, despite the considerable attention given to research on strategy training (Oxford, 2016), the results of this research have been inconclusive (Ardasheva et al., 2017; Gunning & Oxford, 2014; Hassan et al., 2005; Plonsky, 2011). Two meta-analysis studies reveal that some of the research in this area found strategy training resulted in a stronger effect on writing and reading (Hassan et al., 2005), while others reported a stronger effect on speaking (Plonsky, 2011). Also, most of these findings agree that the effect of strategy training on listening is weak and probably negligible (Hassan et al., 2005; Plonsky, 2011). Judging from this information and given the paucity of research conducted on the impact of strategy training on speaking so far, it is apparent that more research is required in this area. The proposed study aims to strengthen the results of strategy training in language learning.

In an earlier study in the 80s, O'Malley et al. (1985) investigated the effect of cognitive, metacognitive, and socio-affective strategies training to improve learners' listening, speaking, and vocabulary skills. The results were not entirely successful as learners significantly improved their speaking performance but not their listening and vocabulary. During the 90s, Cohen et al. (1996) investigated the effect of metacognitive strategies training for oral communication. They trained learners with preparation, self-monitoring, and self-evaluation strategies for conversational speaking tasks. The results revealed that not all of the learners showed improvement in their oral communication skills. Chamot (2005) argues that the unfortunate result of the earlier research on strategies training could be due to a lack of time for the adequate practice of the strategies.

One conclusive result was achieved by Nakatani (2005), who investigated the effects of oral communication strategies (socio-affective) on learners' oral proficiency. This study also provided more practice time for learners. During a 12-week course, learners were trained with metacognitive strategies focused on oral communication strategies such as help-seeking, time-gaining, and negotiation of meaning. The training successfully improved learners' oral proficiency with the dominant use of specific strategies such as maintaining fluency and negotiation of meaning. Recently, Sato (2020) also found the effects of socio-affective strategy training on L2 oral performance. Sato investigated the impact of explicit metacognitive instruction for collaborative interaction (MICI) during communicative tasks. He also compared metacognitive instruction with implicit task-based instruction. The study revealed that learners who received strategy-based instruction outperformed learners who received implicit task-based instruction in both the use of the targeted strategies and the comprehensibility of their discourse.

3.3.1 Fluency Strategy Training

It is argued that learners can be taught specific strategies to improve their oral proficiency in classroom settings (Cohen et al., 1996; de Jong & Perfetti, 2011; Tavakoli et al., 2016). As one critical feature of oral proficiency, speech fluency might also be improved through strategy training instruction (de Jong & Perfetti, 2011; Seifoori & Vahidi, 2012; Tavakoli et

al., 2016). Since speech fluency involves learners' mental process in employing the speech production mechanisms, fluency strategies can be considered cognitive strategies. Brown (2006) proposes that cognitive strategies might be limited to specific learning tasks and involve direct manipulation of the learning material. Fluency strategies training should then be explicitly designed to provide learners with the opportunity to learn and practice specific strategies targeted to improve their speech fluency through classroom instruction. Also, Cohen (2003) suggests that strategy training could cultivate awareness-raising activities and provide multiple opportunities for learners to practice the strategies.

Fluency strategies are defined based on aspects of utterance fluency (Segalowitz, 2010), which consist of speed fluency, breakdown fluency, and repair fluency (de Jong et al., 2013) (see Chapter 4 for more detail). Lambert et al. (2017) also argue that learners can achieve speech fluency by employing speech production mechanisms efficiently and handling performance breakdowns skillfully. A speech that is produced without repairs, hesitations, reformulation, false start, and pauses (either filled or unfilled) might be considered more fluent than one manifesting these features. Therefore, fluency strategies are designed by referring to these features. Fluency strategy training includes paying attention to personal patterns of frequent pausing, effective use of single-word and lexical-chunk fillers, and avoiding unnecessary repair moves (Seifoori & Vahidi, 2012; Tavakoli et al., 2016).

However, research in this specific fluency training is still scarce (Tavakoli et al., 2016) and far from comprehensive. So far, only two studies known to the author, including awareness-raising and multiple practice opportunities, have attempted to train learners to use specific fluency enhancement strategies in a classroom context. The first study was conducted by Seifoori and Vahidi (2012). Throughout 16 grammar-based conversation sessions (one semester), L2 learners were trained to use pre-task planning opportunities and specific fluency enhancement strategies such as using lexical fillers and avoiding disfluency features like repetition and false start to improve learners' fluency and accuracy. Here, the fluency strategies training was conducted in the form of 30-minute awareness-raising activities and task practice. The learners were also exposed to form-focused instruction to improve the correct use of different tenses in the study. The results show that the trained online

planners had an improved speech fluency but with limited evidence of improvement in their accuracy. However, as Tavakoli et al. (2016) pointed out, this study contained three methodical limitations that should be addressed. First, the study's primary aim was to train learners to be more effective in task planning rather than assessing the impact of fluency strategy training per se. Second, the fluency instruction used was embedded in the form-focused context concerned with the accuracy of learners' production. Third, the fluency strategies training was immersed in a planning time component in which learners were taught the strategies while planning for the task performance. Therefore, it is difficult to conclude whether the results were caused by the effects of planning time, proceduralization of the language rules, or development of learner awareness of different aspects of fluency strategies they learned.

Another study by Tavakoli et al. (2016) also attempted to train learners with specific fluency strategy training to rectify some of the issues with Seifoori and Vahidi (2012). Tavakoli and colleagues investigated the effects of pedagogic intervention in fluency strategy training on the development of L2 fluency among learners studying English for academic purposes at a university in the UK. Over four weeks, 37 learners (19 in the experimental group and 18 in the control group) performed a range of monologic tasks. The experimental group received awareness-raising activities and fluency strategy training on all three aspects of utterance fluency: speed, breakdown, and repair. The strategies were taught, including paying attention to personal patterns of frequent pausing, effective use of single-word and lexical-chunk fillers, and avoiding unnecessary repair moves. Learners were given opportunities to practice the learned strategies in the classroom through task performance. The task was retelling a picture story. They were also given homework on retelling another picture story at home. Meanwhile, the control group only received general speaking and listening practice instruction. The results indicated that the experimental group produced statistically more fluent language, demonstrating a faster speech and articulation rate, longer runs, and higher phonation time ratios. The finding revealed an impact of strategy training instruction on the development of learners’ fluency.

However, Tavakoli et al.'s (2016) study was conducted in a specially designed classroom involving L2 learners with mostly intermediate levels of English proficiency. The learners also had exposure to the target language in the UK. It seemed that the strategy training designed to improve learners' speech fluency was effective for this group of learners. However, whether similar strategy training would also be effective for university learners with different backgrounds and contexts has yet to be concluded. For this purpose, classroom-based studies in EFL contexts are required. Even at the university level, EFL learners do not have the exposure to the target language that the ESL learner in Tavakoli et al. (2016)'s study had. Their English proficiency would also be lower, and they usually have a lack of motivation in learning. Since this EFL learning context constitutes a large population of English learners worldwide, a study involving these learners would be beneficial for understanding the impact of strategy training.

3.4 Summary

This chapter has discussed language learning strategy and training as a means to improve L2 speech fluency. A large body of research has attempted to investigate learning strategy and revealed the benefits for learning purposes, including language learning. The application of the three types of learning strategies (metacognitive, cognitive, and socio-affective strategies) in research is often combined with awareness-raising activity in strategy training. Previous research has confirmed the benefit of learning strategy training in improving L2 speaking, writing, listening, and reading. However, research that specifically attempted to investigate the benefits of fluency strategy training is still scarce and less comprehensive. Therefore, more research in this specific context is needed to understand the real benefits of fluency strategy training in improving L2 speech fluency. The following chapter discusses speech fluency measures usually used in research regarding speech fluency.

Chapter 4: Measuring Speech Fluency

This chapter discusses the standard measures used in the research on L2 speech fluency. The discussion begins with Section 4.1, which focuses on measures of utterance fluency, including the temporal measures of speech's quantity, rate (speed), disruption, and repairs. The following section, Section 4.2, highlights the use of these speech fluency measures in previous research. Measures of proceduralization as the indication of automaticity are also discussed in this section. This chapter closes with a table summarizing all the frequently used measures of L2 utterance fluency.

4.1 Utterance Fluency

Segalowitz (2010) classifies fluency into three categories: cognitive fluency, utterance fluency, and perceived fluency. Cognitive fluency is a speaker's *ability* to plan and produce speech effectively. Utterance fluency is the reflection of cognitive fluency speech behaviors. Meanwhile, perceived fluency refers to listeners' impressions of the fluency of a sample of speech. This view considers cognitive sources of L2 fluency as the critical feature of speech fluency. However, since this ability cannot be measured directly, researchers use measures of utterance fluency to gauge speech production process, including speech-planning difficulties in utterances, by counting the number of filled pauses, corrections, and repairs, and by measuring the duration of pauses (de Jong, Groenhout, Schoonen, & Hulstijn, 2015). Hence, utterance fluency is widely applied to research due to its observable and measurable characteristics.

Utterance fluency has been studied by many researchers (Tavakoli et al., 2016) in order to assemble a set of variables that functions as a valid and reliable indicator of speech fluency (Lennon, 1990). Utterance fluency can be defined by objectively measuring (temporal) aspects of a sample speech (Bosker, Pinget, Quené, Sanders, & de Jong, 2013). To be specific, Tavakoli and Skehan (2005) and Skehan (2014a) distinguish utterance fluency into three aspects: speed fluency, breakdown fluency, and repair fluency. Speed fluency is the speed with which a speech is delivered (Götz, 2013). Breakdown fluency is an aspect of fluency that is connected with the continuing flow of speech (de Jong et al., 2012). Repair

fluency is defined as the frequency of overt repairs or reformulations produced by a speaker in making an oral production (de Jong et al., 2012; Götz, 2013).

4.2 Temporal Measures of Utterance Fluency

Utterance fluency designates the temporal variables of speech or the "oral features of utterances that reflect the operation of underlying cognitive processes" (Segalowitz, 2010, p. 48). Previous studies have examined the temporal variables or measures which best predict speech fluency (Bosker et al., 2013; de Jong & Perfetti, 2011; Derwing, Munro, Thomson, & Rossiter, 2009; Foster & Skehan, 1999; Freed, 2000; Kormos & Dénes, 2004; Lennon, 1990; Riggenbach, 1991; Tavakoli et al., 2016; Towell et al., 1996). Ginther, Dimova, and Yang (2010) conclude that the most commonly employed measures of fluency are based on measures of quantity (time) of speech, rate of speech, and disruption in the production of speech. Apart from the measures of quantity, this classification is similar to aspects of utterance fluency suggested by Tavakoli and Skehan (2005). However, since repair measures are not included, the discussion regarding these measures is added below.

4.2.1 Measures of Quantity of Speech

Quantity of production refers to the time spent responding or the number of units (syllables, morphemes, words, and phrases) produced. Total response time, speech time, and phonation time or speech time ratio are three standard measures of quantity. Total response time is the total time required to produce a speech sample, including the time used for producing meaningful utterances, making non-meaningful sounds (fillers/filled pauses), and the time used for silences (silent pauses). Speech time is the amount of time spent producing 'meaningful' utterances without fillers and pauses. Speech time ratio, also known as phonation time ratio, is calculated as the percentage of time spent speaking as a proportion of the total time taken to produce the speech sample (de Jong & Perfetti, 2011). This measure is related to the number of pauses in a speech. For example, when comparing two speech samples like pre-test and post-test, if the mean length of pauses in the post-test is stable, but the number of pauses decreases, phonation time ratio will increase. This ratio is generally assumed as an indication of optimal or problematic speech processing (Kormos

& Dénes, 2004). For example, if a speaker spends less time speaking and more time pausing, the speech time ratio is small. This condition indicates that the speaker has difficulty in formulating the message (Riggenbach, 1991). In measuring speech fluency between pre-test and post-test, increasing the number or percentage of speech time and phonation time ratio represents an increase in speech fluency. Meanwhile, the number of total response times may serve as the basis for calculating speech time and phonation time ratio.

4.2.2 Measures of Rate of Speech

The rate of speech measures refers to the units of speech produced per second or minute. Usually, these measures are based on the number of syllables or words a speaker produces divided by the number of seconds or minutes spent to produce it. The most common measures of rate of production are speech rate, articulation rate, and mean length of runs. Speech rate is measured as the number of syllables divided by total time, including silent pausing time, and therefore incorporates speed of speech and pausing in a speech at the same time (de Jong et al., 2015). This measure is argued to represent all three stages of Levelt's (1989) speech processing model: conceptualization, formulation, and articulation (Towell et al., 1996) (see Chapters 2 and 3 for the detail). Articulation rate is the amount of time required for a speaker to physically produce speech, excluding pauses either filled or unfilled. This measure is argued to represent the efficiency of the articulator (Ginther et al., 2010). Meanwhile, the mean length of run is counted by the mean number of syllables produced divided by the number of pauses in a speech sample, excluding pauses time (de Jong et al., 2013). An increase in speech rate, articulation rate, or mean length of run represents an increase in speech fluency.

4.2.3 Measures of Disruption

Disruptions in speech production represent pauses in speech, including the length, type, and location of pauses. Concerning Levelt's (1989, 1999) speech production model, it is argued that pausing, which refers to silent processing, is the result of serial processing of the stages of speech production (Levelt, 1989). Less proficient L2 speakers are most likely to experience this serial processing because they usually need to apply much thought and

effort to formulate their messages and produce speech (Schmidt, 1992). As a result, the speech is accompanied by pauses and hesitation (Bygate, 2001). In addition, Bygate also argues that pauses happen because the processes of conceptualization, formulation, and articulation of the speech production can occur due to a lack of planning and implementation in less proficient L2 speakers.

Chafe (1980) argues that some speakers pause to make decisions about what to talk about next. Usually, these speakers tend to pause between phrases or clauses. Meanwhile, other speakers pause because of difficulty in deciding how to verbalize the intended message. For these speakers, the pauses tend to fall within phrases and clauses. Usually, less proficient L2 speakers tend to pause within phrases or clauses to allow time for word search.

On the other hand, proficient L2 speakers usually pause between phrases or clauses because they want to convey a complex idea (Towell et al., 1996). It is argued then that when L2 speakers' proficiency increases, they tend to pause more at clause boundaries or between clauses (Lennon, 1990; Towell et al., 1996). Pauses between clauses are argued to be indicators of conceptualization and content planning (Butterworth, 1975). Meanwhile, pauses within clauses indicate breakdowns in the lexical and syntactic encoding (Butterworth, 1975).

Temporal measures of pausing are classified into two: silent and filled pauses. Typically, variation in these forms of pauses at the group level is comparable as L2 speakers tend to be different in the way they pause. Some of them pause in silence while others pause by filling their silence with non-lexical words like *err* and *mm.* Usually, speech fluency increases when silence decreases. In addition to silent pauses, mean length of silent pauses can also be calculated. The idea is to measure any changes in average pause duration from one data point to the other. A shorter average pause duration might also indicate an increase in speech fluency.

4.2.4 Measures of Repair

Repair measures in L2 speech production reflect learners' attention to their errors during the speech. Kormos (1999) argues that the acts of speakers in repairing their errors indicate

a monitoring process during speech production. The existence of repairs in a speech also indicates that the speech production process is "not yet fully automatized," as in the case of less proficient L2 speakers. In other words, advanced L2 speakers have developed a degree of automaticity, so they may potentially create fewer errors that need to be repaired. Advanced L2 speakers with automaticity possess more lexical entries and know grammar rules (declarative knowledge) that they can apply more efficiently to avoid making errors (Kormos, 1999). In addition, advanced L2 speakers have also developed a degree of prodecuralization that allows them to gradually replace the consciously controlled declarative knowledge with the automatic, unconscious rule or memory-based procedures (de Keyser, 1997; Kormos, 1999). Memory-based procedures enable advanced L2 speakers to produce error-free speech (de Keyser, 1997; Robinson, 1997). Hence, speakers' repair phenomena may shed light on automaticity in the speech production process (Kormos, 1999).

Common repair measures in research applications are repetitions, either partial or complete, hesitations, false starts, and reformulations (Tavakoli et al., 2016; Tavakoli & Skehan, 2005). Repetitions are interpreted as the number of words and phrases repeated partially or entirely by a speaker. Repetitions occur when speakers repeat words or phrases in their speech (Tree, 1995). Levelt (1983) argues that repetitions are the signs of speakers' covert editing during speech plans, which occur in the conceptualization stage of speech production. Hesitations signal speakers' difficulties in finding the exact words or phrases to convey meanings. Concerning the speech production process, hesitations may indicate learners' problems during the formulation stage. The appearances of non-lexical fillers such as *err* and *umm* are an indication of hesitations. Measures of filled pauses often replace hesitations measures because they share similarities in characteristics, and both measure disfluency of speech. False starts are the number of abandoned or unfinished clauses. The indications are when speakers start to say something but then decide to abort their utterances and begin again (Tree, 1995). False starts could also indicate speakers' problems during the conceptualization stage. Meanwhile, reformulations are repair moves made by a speaker to change the grammatical rules of clauses or sentences with the new ones.

Reformulations tend to alter the syntax of the repair item (van Wijk & Kempen, 1987), and therefore they could indicate problems in the formulation stage of speakers' speech production process.

4.3 Research on Utterance Fluency

According to Kormos and Dénes (2004), most studies which are concerned with establishing the appropriate measures of speech fluency apply one of the three different approaches: a longitudinal study to investigate the development of fluency (e.g., Derwing et al., 2009; Freed, 2000; Mora & Valls-Ferrer, 2012; Towell et al., 1996), comparison study of fluent and non-fluent speakers (e.g., Riazantseva, 2001; Riggenbach, 1991; Tavakoli, 2010; Tonkyn, 2001), and correlational study relating fluency scores with temporal variables (e.g., Bosker et al., 2013; Derwing et al., 2009; Fulcher, 1996; Kormos & Dénes, 2004; Préfontaine, Kormos, & Johnson, 2016). Most of the studies reported that speech rate and mean length of runs were strongly associated with the development of L2 speech fluency and perceived fluency (e.g., Kormos & Denes, 2004; Lennon, 1990; Towell et al., 1996). Cucchiarini, Strik, and Boves (2002) explored the relationship between objective properties of speech and perceived fluency in reading and spontaneous speech of L2 Dutch learners. The study found that the fluency ratings were highly correlated with speech rate and mean length of runs but not with articulation rate and repair measures. Kormos and Dénes (2004) investigated temporal features of speech produced by intermediate and advanced English learners. They found a significant difference between the two groups of learners in terms of speech rate and mean length of runs. However, their articulation rate and dysfluency measures such as repetitions, restarts, and repairs were not significantly different. The study also found that speech rate and mean length of runs were strongly correlated with fluency scores, while articulation rate and the number of dysfluencies were not. Derwing, Rossiter, Munro, and Thomson (2004) also found associations between temporal variables and L2 perceived fluency with L2 learners. Derwing et al. (2004) used the measures of speech rate, mean length of runs, and pause frequency, and fluency judgments of untrained raters. The study found a significant correlation between perceived fluency ratings, speech rate and pause frequency, accounting for between 65% and 69% of the variance across tasks.

On the other hand, studies have found different results in terms of articulation rate and repair measures. For example, Towell et al. (1996) found that advanced learners of French improved articulation rate after a year at the target language country. Ginther et al. (2010), who investigated relationships between oral English proficiency and temporal measures of fluency, found that articulation rate was strongly correlated with speaking scores. de Jong et al. (2013) explored the effectiveness of correcting measures of L2 fluency for L1 performance to predict L2 proficiency. They found that an inversion model of articulation rate (mean syllable duration) explained 30% of the variance of L2 proficiency. When the measure was corrected for L1 behavior, the explained variance in their study increased to 41%. Bosker et al. (2013) investigated the relative contributions of speed, pauses, and repairs to perceived fluency by examining perceptual sensitivity to the three fluency aspects and found that repairs did contribute a small but significant amount to perceived fluency.

Préfontaine et al. (2016) applied four fluency measures that span only breakdown and speed fluency aspects in their study of L2 French learners. It was found that mean length of run, articulation rate, and length of pauses (average pause time) were the important variables in predicting L2 French fluency perceptions. It can be concluded then that the role of articulation rate and repair measures as indicators of L2 speech fluency is inconclusive.

Research findings regarding pauses also show indecisive results. Two studies by Ginther et al. (2010) and Bosker et al. (2013) found that number and length of pause measures were negatively correlated with proficiency scores and fluency ratings. In contrast, Kormos and Dénes (2004) reported that the length of pauses correlated with fluency ratings while number of pauses showed no correlation. However, in their study, Cucchiarini et al. (2002) found a correlation between fluency ratings and number of pauses but not with the length of pauses. Concerning pause location, Riazantseva (2001) found no difference between L1 and L2 speakers in terms of the number of pauses between clauses they produced in the speech sample.

Recent research findings found that some measures of speech fluency are internally related; therefore, they must be chosen carefully in order to avoid overlapping use (Kormos, 2014;

Skehan, 2014a). As such, Witton-Davies (2014) proposes that the *mean duration of silent pauses*, *number of pauses*, pauses location (between and within clauses), *mean length of run, speech and articulation rates, phonation time ratio*, and a selection of repair measures such as *repetitions* and *reformulations* are the most reliable measures of utterance fluency. The current study used all of the measures proposed by Witton-Davies (2014) for measuring participants' speech fluency.

4.3.1 Measuring Proceduralization

Many studies on speech fluency "are primarily concerned with the degree to which formulation has become proceduralized and how such proceduralization might best be represented by measured variables" (Ginther et al., 2010, p. 382). de Jong and Perfetti (2011) theorized that the proceduralization of language knowledge might happen when learners can produce longer fluent runs in speech. However, it is also theorized that longer fluent runs of speech can occur when a speaker takes more time for planning to produce a sentence de Jong and Perfetti (2011). Usually, when this happens, the speaker will also produce a longer pause time at clause boundaries. "Therefore, if the mean length of fluent runs increases while the mean pause length and phonation/time ratio are stable, more silent planning time was not needed, which indicates that encoding and sentence building have been proceduralized" (de Jong and Perfetti, 2011, p. 539). Next, de Jong and Perfetti (2011) and Towell et al. (1996) proposed the use of mean length of fluent runs as an indicator of proceduralization only when used in combination with the mean length of pauses and phonation time ratio.

To test their theory, de Jong and Perfetti (2011) conducted a study involving 24 L2 learners studying English in America. Participants were distributed into three groups, which received three different conditions with differing time pressure. The first group was asked to repeat the same task three times following the 4/3/2 minutes procedure (Maurice, 1983) intended to create pressure time during the oral performance. The second group performed the same procedure but with three different tasks. Meanwhile, the third group was asked to repeat the same task three times, similar to the first group, but they did not follow the 4/3/2-

minute procedure because they were required to repeat the task performance much later. By comparing pre-test and post-test results, the study found that learners who were asked to repeat the same task without pressure time (the third group) had an increase in their mean length of fluent runs while mean length of pauses and phonation/time ratio was stable. This condition indicates that proceduralization has taken place.

Meanwhile, the first group who received the 4/3/2 procedure showed an increase in phonation/time ratio, a decrease in mean length of pauses, and a stable mean length of fluent runs. This condition also indicates that learners were able to proceduralize their linguistic knowledge. Based on these two results, de Jong and Perfetti (2011) argued that these two patterns of learners' productions could be two alternatives for the manifestations of proceduralization. Proceduralization may be indicated by a speaker's longer runs of fluent speech without additional time for pausing or when the speaker produces the same runs of fluent speech with less pause time.

Earlier, Towell et al. (1996) conducted a study in order to investigate L2 learners' linguistic knowledge proceduralization. The study involved 12 L2 learners of French. They were all British university students who spent a year in France as part of their immersion program to study French. The learners were asked to perform the same task before and after the immersion program. The task required them to retell the story of a French film individually. Towell et al. (1996) used four temporal variables to measures proceduralization, which they argued to attribute to an increase in learners' speech fluency. Those measures are speaking rate, phonation/time ratio, articulation rate, and mean length of fluent runs. The results showed that the learners had an increase in speaking rate and mean length of fluent runs, but their phonation/time ratio and articulation rate remained the same. Towell et al. (1996) argued that the increase in mean length of fluent runs was significant because it was mainly attributable to the proceduralization of different kinds of knowledge, including syntax and lexical phrases.

In summary, Table 4.1 lists nine objective measures most commonly used in speech fluency research, together with how to calculate them. The table includes three measures of

proceduralization: mean length of (fluent) runs, phonation/time ratio, and mean duration of silent pauses that should be used in combination.

Table 4.1 Frequently used measures of utterance fluency

Measure	Formula
Speech rate	Number of syllables/total time
Articulation rate	Number of syllables/speaking time
Phonation Time ratio	Speaking time/total time
Mean Length of run	Total speaking time/number of utterances or number of syllables
Number of (silent) pauses	Number of silent pauses/ number of syllables or total time or speaking time
Number of filled pauses	Number of filled pauses/ number of syllables or total time or speaking time
Mean duration of silent pauses	Pausing time/ number of syllables or number of silent pauses
Number of repetitions	Number of repetitions/ number of syllables or total time or speaking time
Number of repairs	Number of repairs and restarts/ number of syllables or total time or speaking time

The present study used the nine speech fluency measures as detailed in Table 4.1 above. The purpose of using these measures was to explore potentially suitable measures for the context of learners in the study. It is believed that English learners come from various backgrounds and contexts; therefore, different approaches are also required. Having more measures would broaden the opportunity to find the ones that suit learners in a specific unresearched context, such as the context in the present study. In the end, the availability of more measures would help improve researchers' and practitioners' understanding regarding which of these measures best reflect EFL learners' speech fluency and its development in specific contexts.

4.4 Summary

Kormos & Denes (2004) point out that “Just as defining fluency is rather problematical, the establishment of the components of fluency is not without difficulty, either” (Kormos & Dénes, 2004, p. 148). Determining valid and reliable measurements that best predict L2 speech fluency can be problematic because of its complex and multifaceted nature (Tavakoli et al., 2016). Previous studies have been conducted to define the appropriate measures of fluency (Kormos & Dénes, 2004). One distinctive fact from all of these studies is that the speech fluency measures relate to utterance fluency, one aspect of speech fluency proposed by Segalowitz (2010). This chapter discussed general measures of speech fluency, also known as temporal measures, frequently used in studies of fluency. Those measures are categorized as measures of quantity (time) of speech, rate of speech, disruption, and repairs. In addition, a combination of these measures is also used to investigate proceduralization, another indication for speech fluency. Finally, those temporal measures of fluency as listed in Table 4.1 above were used in the present study.

Chapter 5: Complex Dynamic Systems Theory

This chapter highlights Complex Dynamic System Theory (CDST) and its application to L2 research. Section 5.1 explains CDST in brief. Section 5.2 discusses language and its acquisition as a complex dynamic system. The discussion includes the basic characteristics of language systems and the use of phase transition analysis to analyze changes in these systems. The third part summarises CDST applications in L2 research, especially in using the phase transition analysis.

5.1 Complex Dynamic System Theory (CDST)

CDST was originally derived from dynamic system theory (DST) (Abraham, 1994), a branch of mathematics dedicated to the study of systems that are ruled by a consistent set of laws over time (Verspoor, de Bot, & Lowie, 2011). The application can be found in differential equations such as geometrical properties of trajectories and long-term behavior of objects. In the physical world, which is filled with various kinds of systems, DST applications can be found in the motion of planets, whether circulation, animal migration, and the spreading of diseases in the population (Mathematics, 2020).

Systems are the main element in CDST. To be specific, CDST focused on understanding systems that are complex and dynamic. Systems can be understood as a group of entities that work together as a whole. Systems consist of sub-systems that interact with each other. The interactions cause the systems to change continuously or discontinuously (Larsen-Freeman, 2011). The continuous change in a system occurs when the system changes gradually towards its stability (van Geert, 2019). The system could also experience a nongradual change, leading to a discontinuity of the current system's development (van Geert, 2019). When a system experiences a discontinuity, "the growing variable jumps from one stage to the next without intermediary points (van Geert, van der Maas, & Savelsbergh, 1999). This means that the development does not follow gradual developmental steps. In order to go from A to D, for example, development with a continuous change should go to B and C first. However, in a discontinuous change, it could jump from B to D and skip C.

Sometimes, a discontinuous change could happen drastically and make the system exceed its current development state (Van Dijk & Van Geert, 2007). “If this discontinuous change involves the reorganization of the system into a higher-level state (emergence), it is called a phase transition (phase shift)” (Baba and Nitta, 2014, p. 4) (see section 5.2.1.8 for further discussion on phase transitions). These continuous or discontinuous changes that happen in a system make it dynamic (Verspoor et al., 2011, p. 8).

Discontinuities or discontinuous changes are an important aspect of a system development. Discontinuity could a signal that a system undergoes self-organization to reach a higher level of development. A system that experiences a discontinuity could experience a developmental jump by shifting from one stage to the next without intermediary points (van Geert et al. 1999). A developmental jump is often characterized by increasing variability in the system development (van Dijk 2004). The condition of an increased variability in a system is called anomalous variance (van Dijk 2004), which often precedes a developmental jump.

Meanwhile, the word ‘complex’ is embedded in the systems because the dynamic changes they experience are often unpredictable in nature. This unpredictability accompanies much larger, more complex, and non-linear systems (Larsen-Freeman, 1997). In addition, complex systems often consist of many components or agents with behaviors that may not be obvious or easily understood due to a constant interaction among them. This condition causes the system to continually change and becomes highly unpredictable or complex. For example, one cannot make an accurate prediction of the weather system because there are many factors of influence, such as the latitude and altitude of a place, ocean current, temperature, etc. As another example, language is also considered a complex, dynamic system (Beckner et al., 2009; Larsen-Freeman, 1997, 2011) because of its complex, non-linear development. A learner's language development is influenced by factors such as the learner’s social-economic condition, the cognitive processes involved, linguistics of the language, and affective factors such as self-esteem and attitudes. In a language system, these factors of influence are considered as the system’s components, agents, or sub-systems.

Due to this dynamic network of interconnected components, complex systems do not be merely represented by the sum of their components. Moreover, the constant and influential interaction among components together with external input trigger the systems' dynamic behavior which might results in a change in the system. This behavior and change cannot be explained by only examining the components or the external factors themselves. As a whole, the components within a system are interacting with each other as well as with the external factors, and exercising force upon all aspects of the system. Because of its innate characteristic in a system, change, either continuous or discontinuous, becomes an important element in a system's dynamic. Therefore, research involving any complexity science may be focused on describing and explaining changes in a system over time (Byrne, 2005). By knowing what make a system changes, how the changes occur, and what the result of the change is, it is hoped that a similar system could be conditioned to replicate the same changes in order to improve the system. Hence, in language learning research under CDST, the term development marked by changes is preferred instead of acquisition (De Bot & Larsen-Freeman, 2011).

5.2 The Nature of Complex Dynamic Systems: a language learning perspective

5.2.1 The basic characteristic of dynamic systems

Language is a complex and dynamic system (Larsen-Freeman, 2006). Language evolves because of many factors of influence. It also consists of many elements which are interconnected. Language use among individuals and its adaptation to local communities could give rise to "the emergence of certain variegated stabilities of language structure, meaning, and pragmatics" (Larsen-Freeman, 2006, p. 592). It is important to know the basic characteristics of a system to understand language as a complex dynamic system better. Larsen-Freeman (1997) and De Bot and Larsen-Freeman (2011) argue that dynamic systems should possess eight basic characteristics of (1) sensitive dependence to initial conditions, (2) complete interconnectedness, (3) non-linearity in development, (4) change through internal reorganization and interaction with the environment, (5) dependence on internal and external resources, (6) constant change, with chaotic variation sometimes, in which the

systems only temporarily settle into “attractor states”, (7) Iteration, and (8) emergence (phase transition).

5.2.1.1 Sensitive dependence on initial conditions

Complex dynamic systems are known to be sensitively dependent on their initial conditions. This concept is similar to what is known as “the butterfly effect,” which is commonly used in Chaos theory, a branch of mathematics used to study complicated behavior or natural occurrences of complex systems. The butterfly effect simply means that a small difference in the beginning condition of a system can have a big impact on the system in the future. One of its applications is found in meteorology with a famous explanation from a meteorologist, Edward Lorenz. Lorenz (1963) explained that a butterfly's wings might create tiny changes in the atmosphere that may alter the path of a tornado. Here, the wings' flap is considered a part of the initial conditions of an interconnected complex web of weather systems that lead to a tornado occurrence. Hence, the term has been used for various functions outside meteorology.

Complex dynamical systems often show unpredictable non-linear behaviors such that small variances in the initial conditions could have deep and divergent effects on the system’s outcomes (Devaney, 2003). In language learning, the butterfly effect may be interpreted as different learning outcomes that could be triggered by minimal differences between learners (De Bot & Larsen-Freeman, 2011). Different learners in a class may achieve different outcomes, although they have been exposed to the same teaching approaches and materials. For research purposes, De Bot and Larsen-Freeman (2011, p. 10) suggested that it is crucial to gain detailed information regarding learners’ initial condition in order to be able to explain differences and similarities in the outcomes. Individual learner’s first language, experience encounter, and amount of practice with the target language, motivation, self-efficacy, and other relevant information could be crucial for the research. Lowie and Verspoor (2015) argued that the researcher should carefully select the sub-systems to be studied when considering these initial conditions because the contextually dependent

nature of complex systems obliges researchers to know how context invariably shapes the phenomena under study.

5.2.1.2 Complete interconnectedness

Since language is regarded as a complex dynamic system (Beckner et al., 2009; Larsen-Freeman, 1997, 2011), many of its subsystems include the lexical system such as phonological system and syntactical system, are interconnected (De Bot & Larsen-Freeman, 2011). One subsystem may influence other subsystems so that changes in one of them may impact the others. However, the connections between systems and subsystems are not necessarily equally strong. Therefore, the impact of changes resulted from their interactions may not be comparable. De Bot and Larsen-Freeman (2011) gave an example to explain the interconnectedness of language systems based on the impact of three variables (or subsystems) of language learning: language contact, attitude/motivation, and the use of strategies, to the development of language proficiency in three different years. Learners' language contact, attitudes and motivation, and use of strategies at Year 1 influence the change in their language proficiency from Year 1 to Year 2. On the other hand, the change in proficiency may also lead to a change in the state and application of the three variables involved. Furthermore, the change in the three variables may have an impact on the ongoing process of language development from Year 2 to Year 3. The interaction of these three variables is only a small part of the complex, interconnected language system.

5.2.1.3 Nonlinearity in development (chaotic variation)

A causality concept, in which specific causes create specific effects, is something common in the natural world. It could mean that a certain amount of effort will lead to an equal amount of results. For example, if a writer invested twice as much time to write, he would probably produce twice as much writing. Complex dynamic systems do not follow this normal causality concept in which the relationship between cause and effect is nonlinear. In fact, there is a chaotic variation in development that makes it difficult to predict. "There is no proportionate effect for a given cause" (De Bot & Larsen-Freeman, 2011, p. 12) in a complex dynamic system environment.

Meara (2004) demonstrates this nonlinearity concept in his study. He conducted a computer simulation program of vocabulary loss to investigate how attrition of lexical skills works. The computer program provided a network of interconnected words which was operated by switching off one word after another. A series of simulations with different 'attrition events' on the networks revealed that the loss of a certain amount of lexical knowledge occurred without extreme consequences for the overall system. On many occasions, the trajectory of loss demonstrated an initial period of great stability, followed by a dramatic fall where a great proportion of information was lost in a relatively short period of time. After that, the system stabilized again. This simulation showed that switching off one word after another does not cause a similarly linear decline of words in the lexicon and thus provided an example of how nonlinear and chaotic development operates.

This nonlinearity and chaos characteristic of complex dynamic systems development is related to the systems' initial condition and interconnectedness (De Bot & Larsen-Freeman, 2011; Thelen, 2005). System development often occurs in many stages in which each stage's initial condition is the precondition for the stage development. "Once the developmental cascade begins at conception, each state in time is the initial condition of the next instant of time" (Thelen, 2005, p. 265). Different initial conditions will lead to different developmental patterns, which will add to the development's nonlinearity. In addition, "the more components interact, the more problematic it becomes to predict how the system will change" (De Bot & Larsen-Freeman, 2011, p. 12). In practice, learners may have physical fatigue, memory overload, or boredom during a semester-long study. These conditions may suddenly alter the trajectories of current development into something unpredictable. Hence, more variation could happen in development.

5.2.1.4 Change through internal reorganization and interaction with the environment

In the CDST perspective, systems' development is marked by changes in the systems. These changes are caused by internal self-organization and interaction with the environment. In language learning development, learners' motivation, self-efficacy, and cognitive skills are examples of the internal components, while parents, teachers, classmates, classrooms, and

the wider environment are considered external factors. First, the states of internal components will shape the developmental changes in one form. Then, interaction among the components will change the shape into another form. Learners with good motivation, self-efficacy, and cognitive skills, for example, will start their learning better than those who do not. Sometimes, the 'good learners' may experience something in their learning journey that drops their motivation, and this condition may trigger a new developmental change. In addition, the good learners may also deal with teachers or classmates they do not prefer to be deal with. This condition may also cause a change in development. Constant interactions and influences of internal components and external factors cause the system to be complex and dynamic and hence to change continuously or discontinuously.

5.2.1.5 Dependence on internal and external resources

Most systems in the natural world are dependent on internal and external resources. Weather systems need heat from the sun, gravitational pull of the moon, and wind pressure to keep operating. The brain needs oxygen and nutrients to working, and the computer system needs a better processor to keep up with more sophisticated programs. In fact, language learning and language use are also equally dependent on resources (De Bot & Larsen-Freeman, 2011; Van Geert, 2008). Some of the internal resources are learners' memory capacity, problem-solving skills, time to learn, health, well-being, conceptual knowledge, and motivational resources. Examples of external resources are the spatial learning environment, time to learn, learning materials, academic support, and teachers. The internal and external resources constantly interact, and both of these resources are interconnected with the systems. For example, conceptual knowledge could be shaped by learning materials and teachers. Learning materials and teachers may also adapt to learners' conceptual knowledge. Together, all of these resources could change learners' proficiency levels. On the other hand, learners' proficiency level could also determine the suitable learning materials, teachers' qualification, and their conceptual knowledge.

5.2.1.6 Constant change (Attractor States)

Dynamic systems will constantly interact with the environment and reorganize themselves to cope with internal changes (De Bot & Larsen-Freeman, 2011). However, these changes are not linear, and sometimes they settle into an "attractor state" (De Bot & Larsen-Freeman, 2011; Larsen-Freeman, 1997; Lowie & Verspoor, 2015; Van Geert, 2008; Verspoor et al., 2011). Once the systems enter the attractor states, they will require a lot of energy to move into another development state.

Attractor states are the states the systems are situated at a particular point in time as a result of the development process in the systems (De Bot & Larsen-Freeman, 2011). Despite their noticeable appearance within a system's development, the real reason for what makes the systems settle into an attractor state is still unclear. It is also challenging to predict attractor states except when they appear in systems. However, a link between minor differences in the systems' initial conditions and different attractor states can be established (De Bot & Larsen-Freeman, 2011).

The first-year class of English students can be used as an example to describe attractor states. The class can be regarded as a dynamic system because it consists of many components or subsystems: the students, teachers, materials, curriculum, and classroom equipment. Normally, the new class will experience an initial period of transition. Each student starts their personal and collective roles and establishes formal or informal rules for the classroom's interactions. All the classroom's components have some role to play inside the classroom. Sometimes, external forces could influence the classroom, like parents' involvement or interferences from the institution's management. The class's internal organization and external influences drive the classroom to change dynamically over time. Then, in time, the class may stabilize into a cohesive group, and a distinct pattern of behavior emerge, which means that the class enters its attractor state. The class will stay in the attractor state for a while since all the components involved are used to the condition. However, a new state might enter, and the class could move away from the attractor if a

substantial force interferes. It could be in the form of a conflict arising among the students, the arrival of a new teacher, or the departure of a few dominant students.

5.2.1.7 Iteration

The development of a system is an iterative process. However, instead of a simple iteration that just involves repeating merely the same action, iteration in the CDST perspective requires the involvement of other components. A hit song commonly listened in the radio is not just the result of the singer's simple repeated practices. A complex interaction between the singer and the musician, the recording technicians, the equipment, and the environment also plays a crucial role in the song's emergence. Therefore, only repeatedly practicing the song will not make a singer able to produce a hit song if the interaction with other elements in the recording process is not supportive.

One form of Iteration is task repetition. From the CDST perspective, it is argued that iteration may provide learners with the repeated practice of the target language and the opportunity to adapt their language resources to mutable temporal and spatial context and construct their own learning path (Larsen-Freeman, 2018). This argument aligns with soft assembly in language learning proposed by Smith and Thelen (1993). Soft assembly means that in learning the target language, learners soft assemble language construction by themselves like a real-time adjustment made by a speaker during a speech to create meaning and realize their intentions within a specific context. “Soft assembly is a real-time process which takes into account both the options to make meaning and their constraints that are available to speakers, the individual’s language using history, the perceived affordances of context, and the exigencies at the moment” (Larsen-Freeman, 2018, p. 314).

Since iteration leads to learner’s soft assembly, this term is preferred in CDST (Larsen-Freeman, 2018). Furthermore, Larsen-Freeman (1997) argues that there is basically no difference between using language and language change. Suppose learners use specific language elements, the status of the elements in the language system changes. This change will lead to the system’s transformation, such as increasing the probability of the elements being selected in the future or creating new meaning from a current language form. De Bot

and Larsen-Freeman (2011) argue that learners may infer the meaning of an unknown word in reading based on the context, “Each time the word is seen, the previous presentation of the word, with all its semantics, conceptual and syntactic information, will be refreshed and changed slightly due to the context it appeared in” (De Bot & Larsen-Freeman, 2011, p. 16).

5.2.1.8 Emergent properties

Emergent properties of a system are regarded as higher-level properties which are resulted from the interaction of low-level components in the system (De Bot & Larsen-Freeman, 2011). To give an example, De Bot and Larsen-Freeman (2011) explain that the car is the emergent property of its interconnected parts. A car is more than just a collection or assembly of its parts; but its existence is dependent on those parts. If a car is dismantled, then the car will no longer exist. Traffic jam is another example because its occurrence is caused by the interaction of many cars (De Bot & Larsen-Freeman, 2011). Pottery, which is created from clay, may also be considered as an emergence. Becoming pottery is not an inherent characteristic of the clay but an emergent property as a result of interaction among its elements such as water, fire, and the clay itself, and the treatment given by the pottery makers. Therefore, emergence is regarded as an assembly product of some sub-systems or components as the result of the system’s continuous or gradual change (Baba and Nitta, 2014). If an emergence arises from a discontinuous change, then it will be called a phase transition.

5.2.2 Phase Transition

Phase transitions are “points of instability and turbulence where old patterns break down and new ones appear” (Lewis, 2000, p. 39). The appearance of these new patterns is abrupt as the result of the cooperation of all system components. Lewis (2000) also adds that phase transition is usually indicated by a period of extreme variability in the development. As shown by van Geert’s (1994) study, new linguistics abilities increase abruptly from very low to very high levels after experiencing a series of instability. Since its introduction in mathematics, a phase transition has also been observed in investigating natural phenomena in biology and physics. Research investigating human behavior and psychology has also been

discussed in terms of phase transitions. The seminal work of Thelen and Smith (1996) that identified phase transition in an infant's motoric development (change) in stepping and reaching was one major example often referred to by many following studies. The study found that all infants suddenly performed alternating stepping modes that dominated their development at some point during the observation. This sudden shift in infants' stepping behavior is considered a phase transition (Baba & Nitta, 2014). Similar phase transitions were also found in other studies investigating behavioral developments of children. Van Dijk, Hunnius, and van Geert (2012) found that children performed phase transition in their eating behavior when consuming solid food. Jansen and Van der Maas (2001) also found evidence of phase transition, or discontinuous change, in children's motoric development. The evidence was found when the children were asked to predict the movement of a balance scale. In addition, Ruhland and Van Geert (1998) found that children's development of function words change suddenly from telegraphic to differentiated speech. This sudden change is regarded as a phase transition. Hence, in the present study, phase transitions are used to indicate the points at which learners experience extreme variability in their speech fluency development and investigate whether learners experience a new high-level of development (developmental shift) after a phase transition occurs. The findings would be useful to reveal the critical role of phase transitions in learners' cognitive development.

To identify phase transitions, most of the studies employed catastrophe theory (Gilmore, 1993; Thom, 1975; Zeeman, 1976). Catastrophe theory provides a set of clear criteria to identify phase transitions when they occur in a system (Baba & Nitta, 2014; Jansen & Van der Maas, 2001). Catastrophe theory is a general mathematical theory used to detect discontinuities. The theory is applied to investigate sudden qualitative reorganizations (critical points) of the pattern of states of a system (Gilmore, 1993). When the critical points degenerate, discontinuities or transitions happen. Catastrophe theory provides a classification of such degenerate critical points in terms of elementary catastrophes, such as the cusp model. The cusp model is able to explain complex linear and nonlinear relationships simultaneously using 'a high order probability density function' to integrate sudden behavioral jumps (Zeeman, 1976). Furthermore, Gilmore (1993) proposed eight

criteria, so-called catastrophe flags, as the means to identify phase transitions: (1) multimodality, (2) inaccessibility, (3) sudden jumps, (4) divergence, (5) hysteresis, (6) divergence of linear response, (7) critical slowing down, and (8) anomalous variance. Gilmore (1993) argued that if a system experiences a phase transition, then catastrophe flags will emerge near the transition.

5.2.2.1 Multimodality

Multimodality refers to the state in which the system in question may take more than one state for a certain period given a change in a control parameter. In their development to walking, infants' may develop more than one kind of movement, such as crawling and rigid stepping. Infants use their steps to move as their preferred mode, but sometimes they chose to crawl to adapt to their movement or the surface's unbalancing. This condition can be considered as multimodality. Children may also perform multimodality in their language development. Before acquiring differentiated speech (speech with proper grammatical functions), children are found to use both telegraphic (speech without proper grammatical functions) and differentiated speech at different times during their development (Ruhland & Van Geert, 1998). This multimodality can indicate a developmental transition between the old and the new rule or strategy (Van Geert & Van Dijk, 2002).

5.2.2.2 Inaccessibility

The inaccessibility condition is correlated with multimodality. It refers to the mode which is inaccessible during the multimodality's occurrence. When infants perform a combination of crawling and walking (multimodality) at one period, it is almost impossible for them to perform other movement modes. During this period, infants will either crawl or walk. They will never perform an intermediate mode to replace crawling or walking. Similarly, it is unlikely for children to use intermediate speech (a combination between telegraphic and differentiated), which differs from a telegraphic and differentiated speech before they completely use differentiated speech.

5.2.2.3 Sudden Jumps

In their development, infants suddenly can walk after performing a combination of crawling and walking for a while. Similarly, children can abruptly use differentiated speech to speak after using both telegraphic and differentiated speech in their speaking development. This abrupt change from one mode of development to another is called a sudden jump (Baba & Nitta, 2014; Gilmore, 1993; Larsen-Freeman, 2006).

5.2.2.4 Divergence

Devaney (2003) explained that dynamic systems are sensitive depending on their initial conditions (see discussion point 5.2.1.1: *Basic characteristic of dynamic systems*, above). Therefore, small variances or influences in the initial conditions could have strong and divergent effects on the system's outcomes. "When a phase transition is approaching, a small push (influence) may make the system shift to another state. The push may not affect the system substantially at first, but it will become a catalyst for a great transition of the system near phase transition" (Baba & Nitta, 2014, p. 6).

5.2.2.5 Hysteresis

Hysteresis occurs when the jump from one mode of development to another occurs at a different value of the normal variable than the jump in the reverse direction. Zeeman (1976) modeled hysteresis by using the contradictory feeling of fear and anger of dog aggression. Fear and anger belong to the different, opposite poles of the dog's aggression. However, if a frightened dog is continuously provoked, it can make it angry and attack. Once it has attacked, it will continue doing so, although its anger may have reduced. Hysteresis can also be seen as a reversed direction of one or more control parameters. Control parameters are factors that induce highly influential changes in the system at a specific point in time. They make the system unstable and fluctuate to a greater extent (Baba & Nitta, p. 5). When this reversed direction occurs, the system shows some trace of the previous attractor state even after phase transition (Baba & Nitta, 2014).

5.2.2.6 Divergence of linear response

If dynamic systems lose their stability when disturbed in the vicinity of a phase transition, the systems experience what is known as the divergence of linear response. This divergence is connected to the system's reaction to disturbances or perturbations, which cause significant changes in the magnitudes and make the systems unstable (van der Maas & Molenaar, 1992).

5.2.2.7 Critical slowing down

Critical slowing down also connects to systems' reaction to any influential factors near a phase transition. Normally, dynamic systems will return to their stable attractor state after being exposed to a perturbation. However, if the perturbation happens when the systems are approaching a phase transition, the systems become unstable and take a longer time to return to the attractor state (van der Maas & Molenaar, 1992).

5.2.2.8 Anomalous variance

Meanwhile, anomalous variance refers to large increases in the variance of behavioral variables as the transition approaches (van der Maas & Molenaar, 1992). Anomalous variance also causes the systems to become unstable. Usually, the anomalous variance is signaled by fluctuations in the systems' developmental trajectories. These fluctuations are regarded as an important phenomenon in understanding dynamic systems' development because they provide clues of how the systems form patterns of development and discover new and different ways to solve problems (Kelso, 1995).

Baba and Nitta (2014) and van der Maas and Molenaar (1992) argue that in identifying a phase transition, one does not require the presence of all the eight flags. However, it demands multiple flags (at least more than one) in order to demonstrate the validity of a phase transition. "The actual methods or techniques for detecting the flags vary depending on the nature of the research object (Baba & Nitta, 2014, p. 7). Multimodality is commonly found in dynamic systems development as the example of infants' stepping development shown by Smith and Thelen (1993) and children's speech development (Ruhland & Van Geert, 1998). Inaccessibility is also a common flag since it relates to the presence of

multimodality. On the other hand, hysteresis is difficult to examine because it is impossible to reverse the direction of control parameters like educational experiences (Smith & Thelen, 1993). Divergence of linear response and critical slowing down is also difficult to operationalize because it is hard to exactly specify and assess a perturbation effect (van der Maas & Molenaar, 1992). In L2 research involving learners or teachers, divergence can be employed and detected in data involving pre-task questionnaires or interview since they could indicate participants' initial condition before doing the task. Finally, sudden jump and anomalous variance could be investigated in any numerical time-series data.

5.2.2.9 Qualitative change in the attractor

Since catastrophe theory is based on mathematical equations, most analyses rely solely on numerical data by employing a series of statistical tests. This could be considered its limitation in acquiring a deep and resourceful understanding of qualitative development within a system (Baba & Nitta, 2014). Therefore, another criterion for phase transition analysis, called a *qualitative change in the attractor*, is employed (Schmidt, Carello, & Turvey, 1990; Smith & Thelen, 1993). "When a certain threshold is reached for a control parameter, the system undergoes a reorganization that qualitatively differentiates the current mode from the previous mode" (Baba & Nitta, 2014, p. 8). Qualitative change usually occurs following a phase transition. The systems that experience a qualitative change will show new organizational patterns that are different from the previous patterns. For example, in their study, Baba and Nitta (2014) found that learners showed qualitatively changed compositions in their writing after the phase transition. Similarly, Larsen-Freeman (2006) also found that her research participants' oral and written productions changed qualitatively in terms of the idea unit they presented within the four-time (four months) of data collections. Thelen and Smith (1996) also discovered some qualitative changes in infants' motoric development when they shift from crawling to walking. Thelen (2005) revealed that infants also showed a qualitative change in their reaching development. In these two studies, it was found that the qualitative change was mainly caused by the intrinsic dynamic of the infant's body.

5.3 CDST Application in Second Language (L2) Research

Based on the nature and characteristics of complex dynamic systems explained above, research using this approach needs to be conducted in the form of a case study involving individual learners over an extended period of time. The longitudinal case study is a preferred method of data collection in CDST (de Bot, 2015) since it can be used to frequently measure the data and reliably capture the dynamic changes over the timescales of relevance to the theory (Verspoor, Lowie, Chan, & Vahtrick, 2017). The CDST approach can help in gaining a deeper understanding regarding L2 development because it can capture the global (macro-genetic development) and the local (micro-genetic) change in a daily, weekly, or monthly basis. In so doing, Dörnyei (2009) suggests that L2 research could be done in the form of a longitudinal case study involving a single subject and be focused on change rather than variables. This individual-level analysis may reveal the intra-individual variation of non-linear developmental trajectories, which potentially differs from the group's pattern. In so doing, Byrne (2002) and Dörnyei (2014) propose to conduct CDST research by using the retrodiction method of analysis known as retrodictive qualitative modeling. This method is similar to the concept of generalizability in qualitative research and provides a more systematic framework to describe dynamism (Byrne, 2002; Dörnyei, 2014). In research practice, retrodiction can be used to trace the real system in language development and then reconstruct elements, interactions, and change processes of the systems (Larsen-Freeman, 2011; Larsen-Freeman & Cameron, 2008).

Hence, much of the previous L2 studies under the CDST approach have used retrodiction analysis involving small participant groups of a case study (e.g., Baba & Nitta, 2014; Dong, 2016; Evans, 2019; Hepford, 2017; Larsen-Freeman, 2006; Ruhland & Van Geert, 1998; Spoelman & Verspoor, 2010; Verspoor, Lowie, & Van Dijk, 2008). Larsen-Freeman (2006), for example, examined the emergence of complexity, accuracy, and fluency of oral and written production of five Chinese learners of English. The study employed quantitative analysis commonly used to measured complexity, accuracy, and fluency for oral and written production. The focus of this analysis was to find variability in participants' time-series performances as an indication of the discontinuous change (emergent) in their language

development. In addition, the researcher employed qualitative analysis based on the idea unit, especially to reveal how the use of language changes to result in new performances. Both analysis methods found the emergence in the complexity, accuracy, and fluency of participant's language performance. The researcher argued that the emergence was a product of system adaptation to a changing environment in which the participants' language resources transformed differently through language use. In a more technical study, Dong (2016) examined the development of listening strategy use and listening performance in one Chinese EFL learner over a 40- week time window. Dong's study employed dynamic systems techniques of analysis, including the moving min-max graph, loess smoothing, variability, Monte Carlo technique, spline interpolation, moving window correlation and linear regression. These techniques were used to analyze questionnaires' results regarding learner's language strategy use—the questionnaires were administered biweekly during the study. Introspective learner data via participant diaries were collected and analyzed to provide a more contextually dependent listening development account. The study found that learner's listening strategy use and listening performance showed non-linear developmental trajectories with moments of progress and regress.

Despite the intention to reveal learners' non-linear language development (in terms of progressions and regressions that occur during development), most of the above studies did not focus on examining phase transition by applying a clear criterion as offered by the catastrophe theory. The study of Verspoor et al. (2008), for example, observed phenomena similar to phase transition in their study, although it is not the focus of the study. By examining 18 writing samples from a Dutch learner of English over three years, the study found a three-stage development process in the participant's writing in terms of the average length of the writing. The first period was a stable period in which the participant writing was relatively short but gradually increased. The second one was a flux period marked by changes in the writing samples in which performance fluctuates from time to time. Then, the final period occurred when the average length of the writing was stable at a higher level. These flux periods demonstrated the process of phase transition (Verspoor et al., 2008). Using a similar method, Spoelman and Verspoor (2010) found phase transition in the

development of noun phrases in the writing samples of a Dutch learner studying Finnish as a foreign language. The researchers collected 54 writing samples (texts) from the learner for over three years. They noticed that the noun phrase's complexity experienced a sudden jump between the 44th and 45th texts, or in the third year of the study.

To date, only two studies in the L2 context applied such criteria. Those are the study of Ruhland and Van Geert (1998) and Baba and Nitta (2014). Ruhland and Van Geert (1998) investigated the L1 development of six Dutch-speaking children over a 16-month time window. The study started when the children could only produce one-word utterances and ended when they showed a basic command of syntax of closed-class words like pronouns and articles. By employing three flags of the catastrophe theory, sudden jump, multimodality, and anomalous variance, the study found evidence that the children experience a sudden change from telegraphic to differentiated speech. However, the researchers argued that more evidence and techniques were required to prove the existence of catastrophic change in language development.

In another study conducted by Baba and Nitta (2014), three flags of the catastrophe theory (sudden jump, anomalous variance, and divergence) and qualitative change in the attractor analysis were employed to investigate the development of L2 writing of two students enrolled in an English academic writing course in Japan. In addition, to investigate divergence, the study also collected participants' reflective commentary on their writing throughout the course. The study found evidence of phase transitions from both participants based on the four analyses applied. "These findings imply the existence of "supra" patterns of changes in L2 writing fluency, which may contribute to an understanding of processes of individual writers' development" (Baba & Nitta, 2014, p. 30). The present study tried to apply similar methodologies to explore learners' speech fluency development in L2 oral discourse. It is hoped that finding similar "supra" patterns of changes in different research focus might strengthen Baba & Nitta's findings and help explain the fundamental mechanisms of foreign language development.

5.4 Summary

The present chapter has discussed CDST and its application in L2 research. First, the nature of a complex system was explained, followed by the details regarding the complex systems' basic characteristics. Those characteristics are (1) sensitive dependence to initial conditions, (2) complete interconnectedness, (3) non-linearity in development, (4) change through internal reorganization and interaction with the environment, (5) dependence on internal and external resources, (6) constant change, with chaotic variation sometimes, in which the systems only temporarily settle into "attractor states," (7) Iteration, and (8) emergence (phase transition). The discussion was then continued with the phase transition, which is commonly used in the research investigating changes in systems' development, including language learning and its development. Some criteria originally derived from the chaos (catastrophe) theory are used to identify phase transitions. Those criteria are (1) multimodality, (2) inaccessibility, (3) sudden jumps, (4) divergence, (5) hysteresis, (6) divergence of linear response, (7) critical slowing down, and (8) anomalous variance. These criteria are also known as flags. If a system experiences a phase transition, then catastrophe flags will emerge near the transition. Later, another criterion, qualitative change in the attractor, was added to investigate the phase transition. Among those criteria, five of them, sudden jump, multimodality, divergence, anomalous variance, and qualitative change in the attractor, have been successfully applied in previous studies of learners' L2 development. Hence, to investigate phase transitions, the present study used three visible criteria, sudden jump, anomalous variance, and qualitative change in the attractor.

Chapter 6: Research Methods

In this chapter, the design of the study and the methods used in collecting and analyzing the data are described. First, the proposed research questions are outlined in Section 6.1. Following this, an overview of the research design and information regarding the participants' background information are detailed in Sections 6.2 and 6.3, respectively. Details regarding the material used in the study are in Section 6.4. Section 6.5 describes the procedures used to collect the data. Section 6.6 details how the collected data was analyzed. In addition, this chapter includes a clarification of the ethical issues concerning the research process. A summary is given at the end of this chapter.

6.1 Research Questions

This study was an investigation into the effects of two L2 classroom instructional conditions on developing L2 speech fluency: explicit fluency strategy training and implicit task-based instruction. Five research questions were addressed:

Research Question 1: Does explicit fluency strategy training instruction impact L2 learners' speech fluency?

Research Question 2: Does implicit task-based instruction impact L2 learners' speech fluency?

Research Question 3: How does L2 speech fluency change over ten weeks of instruction?

Research Question 4: Do the two forms of instruction impact L2 learners' general oral proficiency as reflected in teachers' judgments?

6.2 Research Design

This quasi-experimental study employed a two-way repeated measures design. The two factors involved were types of instruction (explicit strategy training and implicit task-based instruction) and testing (pre-test and post-test). The dependent variables were nine speech fluency measures: (1) *speech rate,* (2) *articulation rates,* (3) *phonation time ratio,* (4) *mean length of run,* (5) *number of silent pauses,* (6) *number of filled pauses,* (7) *mean duration of*

silent pauses, (8) *number of repetitions*, and (9) *number of repairs.* In addition, the study explored patterns in individual participants' speech fluency development, using a case study approach. All activities in the study were conducted in a computer lab owned by the university, so each participant had one computer for recording purposes.

6.3 Participants

The participants in this study were 54 English students from two intact classes of an English Department in a third-tier vocational university in Indonesia. All participants were second year students (3rd semester) with ages ranging from 18-20 years old. They had completed *Speaking 1* and *Speaking 2* courses in the first year. However, in many cases, especially during the first year, lecturers had to accommodate students' lack of English skills with *Bahasa Indonesia* (the national Indonesian language), which is the students' second language. The students had also studied other compulsory national subjects delivered in this national language during the first year, contributing to some dependency on language use. Participants also used the 'Minang' language (the local language, which is their mother tongue) in their daily interactions. In classes, only a few used English, while the majority were silent observers. Therefore, speaking English was considered the most difficult skill to acquire.

The participants' Test of English as International Communication (TOEIC) scores ranged from 240 to 755 (*N=54, M=492.41, SD=143.41*). This was roughly equivalent to a range from Pre-intermediate or Basic User (A2) to Intermediate or Independent User (B1) on the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). The students in the A2 category might be expected to be able to "understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment)" (Little, 2006, p. 168). These students might also be expected to engage in simple communication regarding their daily routines and familiar activities. In addition to this capacity, B1 students might also be expected to be able to "describe experiences and events, dreams, hopes, and ambitions and briefly give reasons and explanations for opinions and plans" (Little, 2006, p. 168).

All classes at the institution are formed by the institution based on learners' availability and gender (each class is balanced for gender). The students remain in the same class until they finish their studies. Despite the roughly equal number, both groups were comparable in terms of the level of proficiency, age, and gender distribution. Since only two intact classes were involved, they automatically represented the two groups in the study. One group served as the explicit instruction group (EG, n = 29), while the other was the implicit instruction group (IG, n = 25). The number of participants in each group was roughly equal, following the original composition of the intact classes. Table 6.1 summarizes participants' background information across the two groups.

Table 6.1. Summary of participants' background information

Group	N	Age (Year; *Mean*)	Gender Distribution		TOEIC Score (*Mean*)
			Male	Female	
Explicit Group	29	19.30 (*SD* = 0.86)	4	25	453.27 (*SD* = 143.41)
Implicit Group	25	19.24 (*SD* = 0.78)	6	19	537.80 (*SD* = 135.93)

A preliminary analysis was conducted to find a possible correlation between participants' English proficiency, as represented by their TOEIC score, and some speech fluency measures. As detailed in Table 6.2, correlation coefficients between the variables range from *r* = -0.14 (phonation time ratio) to *r* = 0.00 (mean length of run) and reach *r* = 0.14 (articulation rate). It could thus be concluded that participants' English proficiency does not correlate with improvements (gain) in their speech fluency. In other words, learners' TOEIC scores cannot be used to predict gains in speech fluency.

Table 6.2. Correlation between participants' TOEIC score and speech fluency

Relationship	Pearson's r	95 % CI		N
		Lower	Higher	
TOEIC - Speech Rate	0.09	-0.19	0.35	53
TOEIC - Articulation Rate	0.14	-0.13	0.40	53
TOEIC - Phonation Time Ratio	-0.14	-0.39	0.14	53
TOEIC - Mean Length of Run	0.00	-0.27	0.27	53
TOEIC - Number of Pauses	-0.07	-0.33	0.21	53
TOEIC - Mean Duration of Pauses	0.04	-0.23	0.31	53

Correlation coefficients between learners' motivation construct (Kormos, Kiddle, & Csizér, 2011) and speech fluency, as detailed in Table 6.3, were also weak. The correlations between the ideal L2 self and the speech rate, intrinsic motivation and the number of pauses, and internal orientation and the mean length of run are close to zero (r = -0.01). The highest correlation coefficient value in the analysis was found between anxiety and the phonation time ratio (r = -0.23). However, this can also be considered a weak correlation. Thus, learners' responses to the motivational questionnaire used in the study cannot predict their fluency gains.

Table 6.3 The results of estimate correlation analysis of the motivation scales and the speech fluency gains

Relationship			Speech Rate	Articulation Rate	Phonation Time Ratio	Mean Length of Run	Number of Pauses	Mean Duration of Pauses
Ideal L2-Self	Pearson's r		-0.01	0.07	-0.15	-0.02	-0.04	-0.10
	95 % CI	Lower	-0.28	-0.20	-0.40	-0.29	-0.31	-0.36
		Higher	0.26	0.34	0.13	0.26	0.23	0.17
Intrinsic Motivation	Pearson's r		-0.18	-0.19	-0.01	-0.12	-0.01	-0.06
	95 % CI	Lower	-0.43	-0.44	-0.28	-0.38	-0.28	-0.32
		Higher	0.10	0.09	0.26	0.16	0.26	0.22
Self-Efficacy	Pearson's r		-0.05	-0.19	0.15	0.11	0.15	-0.04
	95 % CI	Lower	-0.32	-0.44	-0.13	-0.16	-0.12	-0.31
		Higher	0.22	0.09	0.40	0.37	0.41	0.24
Parental Encouragement	Pearson's r		-0.03	-0.02	-0.05	0.11	-0.02	-0.21
	95 % CI	Lower	-0.30	-0.29	-0.32	-0.17	-0.29	-0.46
		Higher	0.24	0.25	0.23	0.37	0.25	0.06
Anxiety	Pearson's r		-0.18	-0.05	-0.23	-0.07	-0.16	-0.17
	95 % CI	Lower	-0.43	-0.32	-0.47	-0.34	-0.41	-0.42
		Higher	0.09	0.22	0.05	0.20	0.12	0.10
International Orientation	Pearson's r		-0.01	0.12	-0.22	-0.01	-0.03	-0.17
	95 % CI	Lower	-0.28	-0.16	-0.46	-0.28	-0.30	-0.43
		Higher	0.26	0.37	0.06	0.26	0.24	0.10

Meanwhile, correlation analysis between learners' self-regulation and speech fluency, which is informed in Table 6.4, are also weak. The weakest correlation was found between external motivation and the phonation time ratio (r = 0.02). The strongest one is the

correlation between external motivation and the articulation rate (r = 0.35), although this is still considered low (a weak correlation coefficient). Meanwhile, the correlation coefficients between self-regulation and the six speech fluency measures are weaker than the correlation between external motivation and speech fluency. The values range from r = 0.05, the weakest, to r = -0.21. This indicates that learners' self-regulation is only a weak predictor of learners' fluency gains.

Table 6.4 The results of estimate correlation analysis of the self-regulation scales and the speech fluency gains

Relationship			Speech Rate	Articulation Rate	Phonation Time Ratio	Mean Length of Run	Number of Pauses	Mean Duration of Pauses
External Motivation	Pearson's r		0.29	0.35	-0.02	0.30	0.20	-0.05
	95 % CI	Lower	0.03	0.09	-0.29	0.03	-0.08	-0.32
		Higher	0.52	0.57	0.26	0.52	0.44	0.22
Self-Regulation	Pearson's r		0.06	0.18	-0.12	0.05	0.10	-0.21
	95 % CI	Lower	-0.21	-0.10	-0.38	-0.22	-0.17	-0.46
		Higher	0.33	0.43	0.15	0.32	0.36	0.06

These results indicated that participants' proficiency, motivation, and self-regulation did not correlate with speech fluency development. Although some literatures on L2 fluency have revealed that the development of L2 speech fluency could be influenced by many factors, such as learners' socio-cultural background, level of English proficiency, and motivation, for EFL learners in the context of the current study, English proficiency as represented by the TOEIC scores, motivation, and self-regulation as reported by the questionnaires have no relationship with their speech fluency development.

The role of learners' proficiency level in influencing oral fluency has been widely recognised (de Jong et al., 2012). Motivation is also considered to be one of the biggest influences in terms of affective factors that may affect language acquisition (Ellis, 2004), including speaking (Mozgalina, 2015). Meanwhile, self-regulation has a positive influence on learning outcomes (Pintrich et al., 2000); therefore, a positive influence should also be expected on learning the speaking skills. These three constructs of learners' individual differences could

also influence the learners' fluency development in the current study. However, the results show that this was not the case. Therefore, both instructional conditions investigated in the current study seem to be applicable for learners with various individual differences.

In terms of the study's ethical requirements, the present study had obtained letters of permission from the English Department Head and the Director of the institution. Approval from the Department of Education and Training and Curtin University Humanity Research Ethics Committee for minimal risk ethics clearance was also procured. The researcher informed participants that the purpose of the study was to improve the teaching and learning practice for speaking subjects in their department. The researcher acted as a co-instructor as well as a data collector for the study. As this classroom-based study was conducted within the context of regular classes at the institution, participants in the implicit instruction group also received the strategy training material used for the explicit instruction in a form suitable for self-study after the experiment finished. Therefore, all participants received the same instructional material, and comparability between both groups could be maintained.

All participants were asked to sign informed consent regarding their participation in the study. They were informed that their participation was entirely voluntary and that they were free to withdraw from the study at any time without penalty.

6.4 Materials

The materials for the study were pedagogic tasks, questionnaires, an English proficiency test (TOEIC), and oral proficiency rubrics. Three pedagogic tasks were utilized in the study. The main task used to elicit participants' oral speech data involved a monologic news report. Participants had to summarize and report a YouTube news video to radio audiences in their university. This task was output-based, labelled as Task 3. Before performing this monologic news report task, participants also performed two preliminary tasks (labeled as Task 1 and Task 2), which were intended to prepare them for the news report task. Task 1 was input-based and called *identifying factual errors in a news report*. Task 2 was output-based and

named *retelling a news report to a friend.* A complete description of the three tasks undertaken by learners in the present study is provided in Appendix 2.

During the eight instructional meetings in the study, the participants had to perform eight monologic news report tasks (see the procedure section for the detail). The news reports were based on eight YouTube news videos reporting the same genre of news. These videos were dealing with (1) *a terrorist incident,* (2) *a road accident,* (3) *a natural disaster,* (4) *an airplane accident,* (5) *a forest fire,* (6) *an accident at sea,* (7) *domestic violence, and* (8) *hate crime.*

A monologic news report task was also used for the pre-test and post-test. The stories for the pre-test and post-test were different from those learners summarized in class, but similar in length and language difficulty. The videos were: *a terrorist event in Indonesia* for the pre-test and *a natural disaster in Indonesia* for the post-test.

This study also utilized two questionnaires to gather information regarding participants' motivation and self-regulation. All questions in the questionnaires were in the form of five-point Likert scales (Likert, 1932). Participants could only respond to the questions by choosing one of five options prepared for them (see Appendix 1 for the questionnaires' description). The language learners' motivation questionnaire was adopted directly from a previous questionnaire in the literature (Kormos et al., 2011). This consisted of 66 statements under 13 sub-constructs. The Cronbach's alpha reliability of the scales in this questionnaire ranged from 0.71 to 0.88. This motivation questionnaire also contained questions regarding participants' backgrounds and previous learning experiences.

Following the original questionnaire from Kormos et al. (2011), this study also clustered the items into 13 groups (latent variables) and tested for reliability by determining their Cronbach's alpha value (Cronbach, 1951). The first latent variable was named *ideal L2 self,* comprising five items concerning participants' view of themselves as successful L2 speakers (for example, *I can imagine myself reading books and articles in English*). The reliability analysis of this cluster resulted in a Cronbach's alpha value of 0.68, which was below the acceptable value of 0.70 (Field, 2017). The analysis also suggested that the alpha value could

be increased if one item, Question No. 37 (*When I imagine my future job, I see myself using English*), was omitted from the cluster. Therefore, this was deleted, and the new Cronbach's alpha value was 0.80.

The second latent variable was the *intrinsic motivation* which consisted of seven items (for example, *I really enjoy learning English*). The Cronbach's alpha value of this variable was 0.85 and was therefore reliable. The third latent variable consisted of six items relating to participants' belief and confidence about themselves and their English learning (for example, *I am certain that I will be able to use English successfully in my future job*). This variable was called *self-efficacy* with a Cronbach's alpha of 0.80. The fourth variable was *peer pressure,* which contained four items (for example, *My friends are not bothered to study English*). However, its Cronbach's alpha value was only 0.64, and no particular item deletion could improve the value. Therefore, this latent variable was not included in the further correlation analysis.

The fifth latent variable was *parental encouragement,* which contained five items (for example, *My parents really encourage me to study English*). The Cronbach's alpha value of this variable was 0.91. The sixth variable, *instrumental motivation,* consisted of five items that did not result in a suitable Cronbach's alpha value ($\alpha = 0.35$) and this was also excluded from the correlation analysis. The seventh latent variable was *anxiety* and contained eight items (for example, *I get nervous when I'm speaking in my English class*), which resulted in a Cronbach's alpha of 0.82. The eighth latent variable, *technology*, deals with participants' technology usage, and contained three items (for example, *I often use the internet to practice English*). This variable resulted in a very low Cronbach's alpha value ($\alpha = 0.29$) which could not be improved so it also was removed from the correlation analysis.

Latent variable number nine was *resourcefulness,* containing statements regarding participants' use of resources to learn English (for example, *If there is something that I do not understand in the English class, I make efforts to find out more about it*). The Cronbach alpha value was also low ($\alpha = 0.37$) and thus it was omitted from the correlation analysis. The lowest Cronbach's alpha value ($\alpha = -0.02$) was found on the tenth latent variable,

satiation control, which was not included in the correlation analysis. This variable consisted of two items that dealt with participants' ability to overcome a problem in their English learning (for example, *I have my special techniques to make learning English interesting*).

Self-regulation was the eleventh latent variable. This variable contained five items regarding participants' self-regulation ability (for example, *I try to find opportunities to practice English*). As the Cronbach's alpha value was 0.69, it was not included in the correlation analysis. Besides, self-regulation matters were already included in a specific self-regulation questionnaire (see Table 6.3) as discussed later in this section. The twelfth latent variable was *motivational intensity,* which consisted of five items regarding participants' efforts and persistence in learning English (for example, *I'm ready to work hard to learn English*). Its Cronbach's alpha value was low ($\alpha = 0.32$) and, therefore, it was not included in the correlation analysis. The last latent variable was *international orientation* which dealt with participants' attitudes to English as an international language (for example, *I would really like to communicate with foreigners in the future*). The Cronbach's alpha of this variable was 0.72 and therefore was acceptable in terms of reliability and validity.

In summary, out of the 13 latent variables suggested by Kormos et al. (2011), only six resulted in acceptable Cronbach's alpha values ($\alpha = 0.70$ or above). These variables were *ideal L2 self, intrinsic motivation, self-efficacy, parental encouragement, anxiety,* and *international orientation.* The descriptive statistics and reliability analysis results of these latent variables are described in Table 6.2.

Table 6.5. Descriptive statistics and reliability analysis of the motivation questionnaire

Latent variables	Number of items	Mean	Std. deviation	Cronbach's alpha
Ideal_L2_self	4	1.92	0.68	0.80
Intrinsic Motivation	7	1.56	0.54	0.85
Self-Efficacy	6	2.13	0.61	0.80
Parental Encouragement	5	2.24	0.93	0.91
Anxiety	8	2.45	0.67	0.82
International Orientation	7	1.38	0.40	0.72

The self-regulation questionnaire was adopted from Seker (2016). With 30 questions, this questionnaire was centred on three sub-constructs of participants' self-regulation. The

average value of Cronbach's alpha reliability for the three sub-constructs was 0.75. This self-regulation questionnaire was mainly based on Boekaerts' (1997) Self-Regulated Learning Model and Oxford's (1990) Strategy Inventory for Language Learning. The original questionnaire adopted for this study took into account five suggested latent variables, each consisting of items with similar themes. These variables were named *external motivation* (4 items), *internal motivation* (5 items), *cognitive strategies* (7 items), *metacognitive strategies* (10 items), and *evaluation* (4 items). The current study performed a reliability analysis of all five clusters to determine their Cronbach's alpha value. The results of the analysis, together with the descriptive statistics, are presented in Table 6.3.

Table 6.6. Descriptive statistics and reliability analysis of the self-regulation questionnaire

Latent variables	Number of items	Mean	Std. deviation	Cronbach's alpha
External Motivation	4	10.48	3.68	0.76
Internal Motivation	5	7.33	2.31	0.65
Cognitive Strategies	7	15.93	2.95	0.55
Metacognitive Strategies	10	19.39	4.24	0.53
Evaluation	4	8.93	2.21	0.52
Self-regulation	5	10.50	2.75	0.692

As shown in Table 6.3, an acceptable Cronbach's alpha value ($\alpha = 0.72$) was only obtained for intrinsic motivation. The other latent variables were not suitable for correlation analysis. Since no variable could be used to measure participants' self-regulation from this questionnaire, the latent variable measuring self-regulation from the motivation questionnaire described above was used. The Cronbach's alpha value of the variable ($\alpha = 0.692$) was very close to 0.70, and thus taken as a suitable value. Therefore, the correlation analysis of self-regulation included only two latent variables: external motivation and self-regulation.

The TOEIC (Lougheed, 2017) was used to estimate learners' English proficiency. This is a paper-based test with two sections: listening and reading. The listening section consists of 100 items that must be completed within 45 minutes. The reading section must be done within 75 minutes, and it also consists of 100 items. The highest possible score on the TOEIC test is 990, and the lowest is 10.

Finally, two oral proficiency rubrics were used in the study: a holistic and an analytic rubric. First, Brown's (2001) holistic rubric of oral proficiency rating scales was used to rate participants' task performance, based on their overall speech delivery, vocabulary, fluency, pronunciation, and structure. Two raters rated participants into one of Brown's (2001) five levels: (1) beginning speaker, (2) developing speaker, (3) competent speaker, (4) accomplished speaker, and (5) advanced speaker (see Appendix 5 for the description of each level). The same raters rated each performance in terms of Choi's (2005) analytic rubric. This rubric analyzes participants' performance based on five categories: (1) pronunciation, (2) discourse, (3) vocabulary, (4) grammar, and (5) complexity (also see Appendix 5 for the description).

6.5 Procedures

6.5.1 Instruction Procedure

During the instructional meeting or the lesson, both experimental groups (the explicit and the implicit instruction groups) received different treatments. The explicit group (EG) received task repetition practice with explicit strategy instruction, while the implicit instruction group (IG) received task repetition practice without any explicit instruction provided. It was chosen to deliver the instruction during the subject of *Public Speaking* because this subject focuses on improving students' speaking skills and hence accommodated the purpose of the present study. The teacher delivered the instruction for both groups in the *Public Speaking* subject. The researcher acted as the teaching assistant and attended all instructional meetings. There were eight instructional meetings provided for both groups, and each meeting covered one different news topic. Each meeting lasted for two hours and was divided into three learning stages. Although both groups were given the same news stories (see Section 6.5 on Material above), their activities differed slightly. Therefore, the lesson framework, as detailed in Table 6.2, was used to make the instruction time equal in both groups.

Table 6.7. Outline of each two-hour lesson

Stage	EG	Duration (minute)		IG
Input-based task stage	Watching an introduction video	30		Watching an introduction video
Task preparation stage	Explicit Training in Fluency Strategies	30	15	Individual unguided planning time
	Individual guided planning	15	30	Pair practice x 3
	Pair practice x 3	30	30	Pair practice x 3
Output-based task stage	Task performance	15		Task performance

6.5.1.1 Input-Based Task Stage

The input-based task stage of the lesson lasted for 30 minutes. This stage was the same for both groups and was intended to give the participants input regarding the lesson topic. The video of the topic being learned was played twice. For the first eight minutes of the lesson, the teacher informed the participants about the topic and told them that they needed to watch and listen to the video carefully. The participants were also asked to get as much information as they could from the video. Then, they viewed it.

After this, the teacher distributed a fact sheet for each participant to begin Task 1 (see Section 6.5 above and Appendix 2 for details). The fact sheet contained statements about the information or facts from the video. Some of these were correct, according to the video, but others were intentionally incorrect. The task required participants to identify and fix the incorrect statements and then watch the video for the second time to confirm the results. The participants were allowed to refer to the fact sheet while watching the video the second time. They were given five minutes after this viewing to complete their fact sheets. These were then checked by conducting a class discussion led by the teacher. Table 6.3 details the outline for this stage.

Table 6.8. Outline for the 30-minute input-based task stage

Activity	Duration (minutes)
Topic information	8
1st viewing of video	2
Task explanation and fact sheet distribution	3
2nd viewing of video	2
Task completion	5
Whole class checking and discussion	10

6.5.1.2 Task Preparation Stage

Unlike the input-based task stage, the two groups went through different treatments during the task preparation stage. The EG started their fluency strategy training while the IG began their task practice. As shown in Table 6.2, the EG had three activities, and the IG only had two. In this stage, 75 minutes were allocated to complete all activities. The following discussions detail the procedures undergone by each group during this task preparation stage.

A. Activities in the EG

The three activities conducted by the EG in this stage were: fluency strategy training, individual guided planning, and practice.

Fluency Strategy Training. The fluency strategy training was adapted from similar fluency strategy training developed by Seifoori and Vahidi (2012) and Tavakoli et al. (2016). The training was focused on utilizing aspects of utterance fluency (speed, breakdown, and repair) which were incorporated into specific strategies such as (1) paying attention to personal patterns of frequent pausing, (2) practical use of single word and lexical chunk fillers, (3) avoiding repetition, (4) avoiding false starts, (5) avoiding reformulation, and (6) avoiding replacement. These strategies were then distributed across the eight instructional meetings (see Table 6.4 for details).

The 30 minutes fluency strategy training designed for the EG consisted of 10 minutes awareness-raising and 20 minutes strategy practice. The awareness-raising activity was intended to make the participants aware of some aspects of utterance fluency, such as pauses, repetition, and repair moves that could influence the fluency of speech. They learned some examples of these utterance fluency aspects from two sources: a YouTube news video presented by a native speaker and an audio recording from a former student of the *Public Speaking* class. The student's recording was a reporting speech of a YouTube news video. The participants were asked to pay attention to aspects of utterance fluency that could be found in the two sources. To be specific, the student's recording was used to show the participants some examples of breakdown fluency because it contained evidence of pauses, repetitions, and repairs. Meanwhile, the YouTube video provided examples of fillers and some repairs. They were also informed that the speakers' speech fluency could be improved if they used the fluency strategies.

In the next 20 minutes, participants were asked to apply the strategy they were made aware of by reproducing the news video they had watched during the input-based task stage. They were explicitly asked to use the strategy they had just learned to make their speech more fluent. Table 6.4 summarizes the strategy associated with each topic.

Table 6.9. List of strategies and news topics in each lesson

Lesson	Strategy	Topic
1	Paying attention to personal patterns of frequent pausing	Terror incident
2	Using single sound and word fillers (well, um/er/uh, hmm, like, basically, actually, etc.)	Road accident
3	Using lexical chunk fillers (you see, you know, I mean, I suppose, I guess, I think, etc.)	Natural disaster
4	Avoiding repetition	Airplane accident
5	Avoiding replacement	Forest fire
6	Avoiding false starts	Accident at sea
7	Avoiding reformulation	Hate crime

Lesson	Strategy	Topic
8	Application all the above strategies	Domestic violence

In the first meeting, the training started with an awareness-raising activity in which the participants were introduced to speech fluency and one specific aspect of utterance fluency. In this meeting, the training focused on pauses as an aspect of breakdown fluency. They learned that frequent pausing affected speech fluency in terms of the speech rate, therefore, it should be avoided. Next, the participants were asked to listen to the same student's recording they had during the awareness-raising. They were explicitly asked to analyze where the speaker's frequent pausing occurred. Then, a class discussion was held to discuss the result of the participants' analysis. After that, the participants were assigned to make a news report based on the same YouTube video and focus on eliminating any possible pauses in their performance.

In the second meeting, the training still focused on breakdown fluency. This time, the participants learned how to fill their pauses to improve their fluency. As their awareness-raising activity, a YouTube news video of a native news anchor was introduced. The participants were asked to pay attention to the way the reporter filled pauses with some single word fillers. Next, a range of single word fillers (*well, like, actually, basically*) was introduced to the participants. They were then asked to make their own report based on the YouTube video and practise using the fillers to fill pauses.

In the third meeting, *lexical chunk fillers* were targeted as the strategy to deal with pauses. An awareness-raising activity also preceded this meeting in which the participants watched a YouTube news video presented by a native news anchor. The participants were asked to pay attention to the lexical chunk fillers used by the presenter. After that, the participants learned about various lexical chunk fillers that could be used in their report. They were then assigned to practise reporting the same news video on their own by using the newly learned fillers when necessary.

The training for *repetition* was held in the fourth meeting. For awareness-raising, the participants were informed that repeated words, phrases, or utterances during a speech affected fluency of speech. After that, they listened to a sample recording from one of the former students to identify any repetitions in the recording. Their findings were then discussed with the entire class. After this, the participants started their practice time. They were asked to repeat the report individually by trying to avoid repetition in their speech.

In the fifth meeting, the participants were asked to focus on avoiding *replacements* or using lexical items that were immediately substituted for another. They listened to a former student's speech recording to find some examples of replacement and discussed their findings together as the awareness-raising activity. For practice, the participants made their own report based on the former students' recording, to apply the strategy.

In the sixth meeting, the training focused on *false starts.* Here, the participants were trained to avoid false starts, which are incomplete or abandoned utterances. Again, they learned and discussed some examples from a former student's recording and then practised producing the same report by avoiding any possible *false starts*.

Reformulation was introduced in the seventh meeting. The participants were informed that reformulation or repeating phrases and clauses with some modification to syntax, morphology, or word order in their speech could also harm their speech fluency, and they were asked to avoid this while speaking. The participants also learned about reformulation from a former student's recording. Next, they practised making the same report by avoiding making any reformulation moves.

Finally, in the eighth meeting, the participants practised all the previously learned strategies. First, they analyzed a sample recording to identify all aspects of utterance fluency in it and discussed their findings. Then, they practised reporting the same news by using all the fluency strategies they had learned since the first meeting. They were explicitly informed to use single word or lexical chunk fillers to fill their potential pauses during their speech and avoid any possible repair moves.

Individual Guided Planning. After the fluency strategy training, the students were given 15 minutes of guided pre-task planning on an individual basis to prepare for performing Task 2 (see Section 6.5 and Appendix 2). They planned to summarize a news story for a friend while focusing particularly on the fluency strategy that had been covered in the lesson on that day. Although previously researchers such as Ellis and Yuan (2004), Foster and Skehan (1996), and Wendel (1997) suggested 10 minutes of pre-task planning, participants in this study were given 15 minutes so they would have enough time to read and understand the text. Considering their proficiency levels, these participants required this amount of time for reading with comprehension. The teacher also supervised participants' planning activity and reminded them to pay attention to the speech fluency aspects being learned. For the next 30 minutes, the participants practised Task 2.

Practice. The practice was designed to allow each student to perform Task 2 three times with three different partners. The participants were provided with two written news stories with a completed fact sheet (see Appendix 2 for Task 2 materials). Each of them was asked to find a partner who had a different news story. The pair then took turns giving and listening to the speech. The speaker needed to provide a 1-minute report to the listener based on the provided news story. The listener who had a fact sheet containing a list of information from the speaker's news story needed to clarify any missing information from the fact sheet and monitor the speaker's speaking time. Once they completed their turns, the participants were then asked to repeat the same activity with two other partners so each of them would perform the same role three times. The instructor monitored the practice activity and reminded participants to apply the strategy that had been learned. The time allocated for all activities at this stage was 30 minutes.

B. Activities in IG

Unlike the EG, the IG only had two activities but with an equal amount of time spent at this task preparation stage. Following the input-based task stage, the participants in this group were not given any training regarding the speech fluency aspects. Instead, they started with a 15-minute unguided pre-task planning, and then continued with the practice activity. The

participants conducted their planning activity entirely by themselves without any involvement of the teacher. The participants' tasks and activities in practice were identical to what was conducted by the EG, but this group completed the cycle twice (2 x 3 practices). This design ensured that both groups spent equal time in the practice stage.

6.5.1.3 Output-based Task Stage

The final stage of the lesson was the same for both groups. Here, the participants performed Task 3, as described in Section 6.5. The participants were asked to perform a 1 minute news report speech based on a YouTube news video. They performed the task three times by recording it on a computer. They were also encouraged to think about it and conduct the task very carefully, as if their report might be broadcast on the university's radio station. This stage began with the students watching the news video twice during which the participants could make notes. They could use the notes for their first two recordings but not for their third one. Therefore, all notes were taken by the teacher before they recorded their speech for the third time. This final recording was used for the analysis.

6.6 Data Analysis

Three forms of research data were collected in the study: audio recordings of participants' oral performances, English proficiency (TOEIC) test, and questionnaires. The audio recordings consisted of eight oral task performances and two oral tests: a pre-test and post-test. All participants performed the eight oral tasks at the end of every meeting in Week 2 to Week 9. The pre-test and post-test data were gathered in Week 1 and Week 10, respectively. In total, 540 audio recordings of approximately 540 minutes' total duration were collected from the 54 participants during the ten-week study.

All participants also completed TOEIC proficiency tests and the questionnaires before the study started. The TOEIC test data consisted of 54 English proficiency scores of the listening and reading. The questionnaire data consisted of participants' responses to questions regarding their motivation and self-regulation. The purpose of these data was to establish participants' initial comparability and provide a basis for examining the moderating effects of motivation and self-efficacy on the impact of instruction.

All the collected data were then prepared for analysis: inferential analysis, estimation analysis, and phase transition analysis (see Section 6.5.4 below for the discussion on the three methods of analysis). The preparation included transcription and coding of 140 audio recordings from all participants' pre-test (54) and post-test (54) recordings and eight task performance recordings from four participants (32 in total) selected for the phase transition analysis.

6.6.1 Data Transcription and Coding

The 140 audio recordings were first transcribed manually by a teacher from the English Department who also acted as a research assistant for the study. The transcriptions were then checked and coded for pauses by the researcher using PRAAT 6.1.09 (Boersma & Weenink, 2019). Meanwhile, the coding for other speech fluency measures was done manually in Microsoft Excel 2016. The spreadsheet also contained information regarding each participant's total time and speaking time. The total time was the time used by speakers to produce a speech, including pauses. The speaking time was the actual time spent speaking, after all the pause time was removed. In other words, the speaking time was the result of subtracting the pause time from the total time.

The mean duration (length) of the pause and phonation time ratios were taken as measures of proceduralization. The mean duration of pauses was the total length of pauses in seconds divided by the total number of pauses. The phonation time ratio was obtained when the speaking time was divided by the total time (see Chapter 4 Section 4.3.1 for details). The third measure of proceduralization was the mean length of run or the mean number of syllables produced between pauses. The number of syllables was counted based on pruned transcripts, using an online syllable counter (Arczis Web Technologies, 2019). These pruned transcripts were created after omitting any repeated words, false starts, reformulated or replaced words or phrases, unrecognized utterances, greetings in speech, and opening and closing speech including the speakers' self-introductions (see Appendix 4 for some examples of greetings, opening and closing speech, and self-introductions made by the participants).

The speech rate and articulation rate were used as the measures of speed fluency. The speech rate was measured as the number of syllables divided by the total time. The articulation rate was calculated as the number of syllables divided by the speaking time. Other variables used to measure speech fluency in the study were verbatim repetitions and repairs. The repair measures included false starts, reformulations, and replacements. These measures were divided by the number of syllables to maintain comparability across speech samples (see Chapter 4 Section 4.2.4 for more details on repair measures).

6.6.1.1 Pauses

Coding for pauses was done with PRAAT. First, the beginning and end of each speech segment was determined by using a function in the software named “To textgrid (silences).” The textgrid was useful to check and adjust pause boundaries because it has a spectrogram and waveform display that could be visually inspected. The inspection was also done by listening carefully to the recording while looking at the textgrid. There were two kinds of pauses coded for the purposes of analysis: unfilled pauses and filled pauses. The unfilled pause was defined as a silence of .25 second or longer (Arczis Web Technologies, 2019; de Jong et al., 2013; Raupach, 1987). Towell et al. (1996) argue that .25 second is a considerable cut-off point because "if the point was too low, the analyst may be confused by displays in which an apparent pause is, in fact, the top phase of gemmated plosives or other normal phenomena. If it is too high, the analyst may omit significant amounts of pause time" (Towell et al., 1996, p. 91).

There were two kinds of unfilled pauses coded in the study. The first was unfilled pauses between clause boundaries, and the second was unfilled pauses within a clause boundary. Filled pauses were all non-verbal fillers such as “uh,” “ah,” “err,” and “mm” (de Jong & Perfetti, 2011). All pauses were coded according to their category and location and their duration was written in seconds, within the transcription in Microsoft Word.

Once the coding for pauses was completed, the results were then moved into the Microsoft Excel spreadsheet. Each utterance was put into one line in the spreadsheet. The pauses in each utterance were then counted manually, and the results were put into columns next to

the utterance. One column represented one category of pause (filled or unfilled pause), and each of these columns was accompanied by a column for storing information regarding the duration of each pause. The total number and the total duration of pauses were calculated at the bottom of each column.

6.6.1.2 Repairs

Next, more columns in the Excel spreadsheet were created to store information for the coding of three repair measures: false starts, replacements, or reformulation. These three repair measures were coded by manually counting them in each utterance. One column was provided for each measure, with totals calculated at the bottom of each column.

6.6.1.3 Measures of Proceduralization

Three measures of proceduralization were calculated using formulas in Excel. First, the mean duration of pauses was generated by calculating all pauses' durations (pause lengths in second). The result of the calculation was then divided by the total number of pauses. For this purpose, all the unfilled and filled pauses were included. Second, the phonation/time ratio (PTR) was computed by dividing the total time filled with speech (not including pauses) by the total time spent speaking (time filled with speech + pauses). Third, the mean length of fluent run was calculated as the mean number of syllables between pauses. The coding was derived from dividing the number of pauses by the number of syllables in the pruned transcription.

6.6.1.4 Other Measures

The three other measures in the study were the verbatim repetitions, speech rate, and articulation rate. Verbatim repetitions were counted by manual inspection of each utterance. The speech rate was calculated by dividing the number of syllables by the total time spent speaking (time filled with speech + pauses). The articulation rate was derived by dividing the number of syllables by the speaking time. Together, nine speech fluency measures were used as the dependent variables in the study (see Table 6.5).

Table 6.10. Speech fluency measures used in the study

No.	Measure	Acronyms
1	Speech Rate	SR
2	Articulation Rate	AR
3	Phonation/Time Ratio	PTR
4	Mean Length of Run	MLR
5	Number of (silent) Pauses	NP
6	Number of Filled Pauses	NFP
7	Mean Duration of (silent) Pauses	MDP
8	Number of Repetitions	Rep
9	Number of Repairs	Repair

6.6.2 Method of Analysis

There were three analysis methods involved in this study: inferential analysis, estimation analysis, and phase transition analysis. The researcher used non-parametric tests of the inferential analysis as the data was found to be highly skewed and not normally distributed. The common data transformational tools, such as logarithmic and square root, failed to improve the data. Hence, the analysis needed to be conducted with non-parametric tests. Mann-Whitney U and Wilcoxon Signed Rank tests were used. The following discussions detail the other two methods which are relatively new in the applied linguistics: estimation analysis and phase transition analysis.

6.6.2.1 The Estimation Statistical Analysis

Estimation statistics are methods of statistical analysis that focus on the estimation of effect sizes or the point of estimates and their confidence intervals (precision estimates) (Claridge-Chang & Assam, 2016; Cumming, 2014). To be specific, estimations statistics explain the magnitude and precision of an effect being investigated. This analysis method was chosen because “knowing and thinking about the magnitude and precision of an effect is more useful to quantitative science than contemplating the probability of observing data of at least that extremity, assuming absolutely no effect” (Claridge-Chang & Assam, 2016, p. 1).

Estimation statistics could offer a solution to a study that fails to theoretically or pedagogically contribute to knowledge because of its non-significant result. This data analysis could preclude a study from the dichotomous thinking of null hypothesis significance testing, which unintentionally requires a study to achieve a significant result. For a classroom-based study, such as the present study, getting a significant result is not guaranteed because of the limited number of participants and potential extreme values or values close to zero or a natural limit in the data. Therefore, focusing a study on exploring the magnitude of the effect being investigated might be better than focusing it on the level of significance.

Estimation has been described as "a comprehensive analysis framework that offers a better way to interpret data, quantitative literature review methods and techniques to analyze heterogeneous data" (Claridge-Chang & Assam, 2016, p. 2). Furthermore, Cumming (2014) argues that data analysis and interpretation using estimation statistics are more informative than the null hypothesis significance testing and *p* values. For example, if a poll report stated that voters prefer Candidate A for 27% with an error margin of 2%, it is understood that the 27% was derived from a sample. Therefore, it may be assumed that 27% is a fair estimate of support for Candidate A in the population. With the 2% being considered as the largest likely error in the poll, the natural and informative way to present the poll result for Candidate A should be 27% ± 2%, or a 95% confidence interval (CI) of (25, 29). The 27% is the point estimate, while the CI is the interval estimate that indicates the precision of the estimation or "the best point estimate of what we want to know" (Cumming, 2014, p. 13). By using a test of significance and the *p*-value, the result, which should be presented by stating that the support is significantly greater than 25% with $p < 0.01$, seems to be less informative for wider audiences.

Effect Sizes (ESs)

According to Cumming and Fidler (2009), an effect size (ES) is an amount of anything of interest in data analysis such as means, differences between means, correlations, and other familiar quantities in statistics. Two kinds of effect sizes exist: the original units (e.g.,

seconds, word numbers, or score units) and the standardized or unit-free measures like Cohen's *d* (Cumming, 2014; Cumming & Fidler, 2009). ESs in the form of the original units are readily interpreted and should be reported accordingly. The standardized units are useful for comparing studies with similar cohorts and for meta-analysis (Cumming, 2014). Due to the benefits offered by both kinds of ESs, both were reported in the present study.

The original ESs were reported in the form of original scores of the speech fluency measures as discussed in Chapter 4 Section 4.2 and 4.3 (see Table 4.1 for the detail). Meanwhile, for the standardized unit, Cohen's *d* (Cohen, 1969) was used. To be specific, in the analysis, Cohen's *d* unbiased (d_{unb}) and *d* average (d_{avg}) were used because "the usual Cohen's *d* only gives values that are a biased estimate of the population effect size" (Cumming, 2014, p. 16). Therefore, a simple adjustment was required to achieve the unbiased version of Cohen's *d.*

In the present study, Cohen's d_{unb} was produced by using ESCI (Exploratory Software for Confidence Intervals) (Cumming, 2020), an open-source program for statistical analysis. In addition, Cohen's *d* interpretation was also used to determine the magnitude of the effect. Hence, $d = 0.2$ is considered a 'small' effect size, $d = 0.5$ represents a 'medium' effect size, and $d = 0.8$ constitutes a large effect size.

Confidence Intervals

A CI is "an estimated range of values with a given high probability of covering the true population value" (Hays, 1973, p. 375). CI is also interpreted as an interval estimate that indicates the precision of a point estimate or ES (Cumming & Finch, 2005). The general convention of a CI is 95% which means that the point estimate has a 95% chance to be similar to the population estimate. The CI in the present study was also derived from the ESCI program.

6.6.2.2 Phase Transition Analysis

Phase transition analysis is a part of complex dynamic systems (CDS) approach, which aims to explain the trajectory of a system, or a developmental path experienced by individual

learners (see the explanation in Chapter 5). This approach emerged because, in some research, "the central tendency observed in a group may not be true of any particular person in the participant sample" (Dörnyei, 2014, p. 83). Since a system trajectory is affected by many factors, CDS does not endeavour to predict system change but rather to "retroactively explain the trajectory by extracting and describing its characteristics or patterns" (Baba & Nitta, 2014, p. 14).

In data analysis, this approach is called retrodiction, which is derived from retrodictive qualitative modeling (RQM) (Dörnyei, 2014). RQM attempts to trace the real system in language development and then reconstruct the elements, interactions, and change processes of the systems (Dörnyei, 2014; Larsen-Freeman, 2011; Larsen-Freeman & Cameron, 2008). The RQM approach can provide an understanding regarding "salient patterns – or essential underlying mechanisms – associated with typical system outcomes" (Dörnyei, 2014, p. 89). This approach is similar to the concept of generalizability in qualitative research and provides a more systematic framework to describe dynamism (Dörnyei, 2014).

Retrodiction analysis was performed by using a case-based method to find a developmental change (phase transition) in a complex system. This method makes a case (a complex system itself) the target rather than a variable (Baba & Nitta, 2014). Byrne and Ragin (2009) argue that a case has trajectories and transformations that depend on the interactions with other parts or systems within or outside the case.

Two participants from each group were selected for the phase transition analysis. The pre-tests, post-tests, and eight oral performance recordings data from four participants were prepared for analysis. All of these participants were those with high fluency gains, as reflected by their gains in speech rate and proceduralization, which include the mean length of runs, mean duration of pauses, and phonation time ratios (see Chapter 4 section 4.3.1 for the details). The four participants' data were analyzed to find signs of phase transitions.

The analysis included the number of syllables instead of the number of words, which was used in Baba and Nitta's (2014) study. The reason for this decision was that most of the

speech fluency measures used in the present study were based on or divided by the number of syllables. For example, the speech rate resulted from dividing the number of syllables by the total time, and the articulation rate resulted from dividing the number of syllables by the speech time. Other measures such as the mean number of pauses, the mean length of run, and the mean duration of pauses were all divided by the number of syllables to get the operational value (see Chapter 4 Section 4.2 and 4.3 for the details regarding the operationalization of the speech fluency measures). The number of syllables could be considered more effective than the number of words in capturing the change in speech development. If learners use words containing more syllables in their speech, then the developmental changes of their number of syllables will be more visible for a period of time.

Following Baba and Nitta (2014), the phase transition analysis was based on three criteria: (1) a sudden jump or an abrupt change in the system from one state to another, (2) anomalous variance or a phenomenon that happens in the system when it becomes unstable and shows an increasing variability as the transition approaches, and (3) a qualitative change in the attractor or a system reorganization which qualitatively differentiates the current mode from the previous one.

The sudden jumps were identified by using change point analysis, indicating the points at which a sudden change occurs in the data (Steenbeek et al., 2012). The change point analysis used calculations derived from Change Point Analyzer software (Taylor, 2000). This analysis was followed by a moving min-max graph (Van Geert & Van Dijk, 2002; Verspoor et al., 2011) to further examine the identified change points. The anomalous variance was analyzed by using a line graph of the moving standard deviation of residuals to visualize the greatness of fluctuations in text length (Van Geert & Van Dijk, 2002). For the qualitative change in the attractor, the pruned transcripts of the task performances were assessed qualitatively to find the indicators.

6.6.2.3 Raters' Judgement on Oral Performance

To evaluate participants' oral proficiency, raters' judgement was used. The two raters were non-native English teachers with extensive experience teaching English to Indonesian

learners. Both were active teachers in the English Department of the University who had no previous teaching contact with any of the participants. To avoid subjectivity in their judgement, participants' recordings were labelled with numbers instead of their names. Two oral proficiency rubrics (see Section 6.5) were used for this rating purpose.

6.7 Summary

In this chapter, the study's research questions were outlined. Also discussed were the background of the participants, the design, and the used materials. Specific procedures used to collect and analyze the data have also been detailed in this chapter. These include the three analysis methods applied in this study: inferential analysis, estimation analysis, and phase transition analysis. This chapter also provided details about how the data were transcribed and coded for analysis. The next chapter consists of a discussion of the results of the estimation statistical analyses used to answer the research questions.

Chapter 7: The Inferential and Estimation Analyses Results

This chapter comprises a discussion of the results of the statistical analyses conducted. In Section 7.1, the tests to screen the data on each of the dependent variables are described. The screening procedures consisted of statistical tests to confirm the normality of score distributions, detect outliers, and confirm the homogeneity of the variance on each of the variables. It was found that the data were not normally distributed or homogeneous. Traditional parametric inferential statistics could thus not be used in the study. Due to the small sample sizes and lack of statistical power, non-parametric statistics were employed to investigate participants' speech fluency and oral proficiency differences within and between the experimental groups. In addition, to unmask a potential meaningful variation in the data, new statistics based on estimation analysis (Cumming, 2014; Plonsky, 2015) were used. The estimation analysis was particularly applied to estimate absolute effects in terms of effect sizes and confidence intervals. In Section 7.2, the descriptive analyses are given for each of the groups separately. The results of the inferential and estimation analyses are then presented in Section 7.3 where Pearson's *r* correlations are also detailed, to evaluate the relationships between variables. A summary of the analyses results is provided in Section 7.4.

7.1 Data Screening

The analyses in the current study employed nine dependent variables: (1) speech rate, (2) articulation rate, (3) phonation time ratio, (4) mean length of run, (5) number of silent pauses, (6) number of filled pauses, (7) repetitions, and (9) repairs. The repairs consisted of three measures: false starts, reformulations, and replacements. First, all variables were screened for normality using the Shapiro-Wilk test ($p < .05$) and the z-scores for skew and kurtosis ($p > .05$). Then, a visual inspection of histograms, normal Q-Q plots, and box plots for each of the variables was conducted. In addition, a series of Levene's F-tests (Martin & Bridgman, 2012) ($p > .05$) was used to identify homogeneity of variance. The results of the data screening are presented in Table 7.1.

Table 7.1 The results of data screening

No	Variable	Normality (*p* for Shapiro-Wilkes)	Z-score		*p* for Levene's test
			Skewness	Kurtosis	
1	Speech rate pre-test	0.35	1.83	1.71	0.07
	Speech rate post-test	0.06	2.46	2.70	0.27
2	Articulation rate pre-test	0.15	1.77	-0.22	0.69
	Articulation rate post-test	0.35	1.10	-0.76	0.15
3	Phonation time ratio pre-test	0.05	-2.36	1.87	0.53
	Phonation time ratio post-test	0.76	0.87	-0.61	0.61
4	Mean length of run pre-test	0.00	7.68	15.48	0.01
	Mean length of run post-test	0.00	2.99	6.64	0.99
5	Number of (silent) pauses pre-test	0.01	2.98	2.23	0.05
	Number of (silent) pauses post-test	0.70	1.47	1.25	0.42
6	Number of filled pauses pre-test	0.00	4.66	3.12	0.64
	Number of filled pauses post-test	0.00	12.05	33.45	0.33
7	Mean duration of pauses pre-test	0.00	4.31	6.22	0.08
	Mean duration of pauses post-test	0.01	2.60	1.26	0.03
8	Number of repetitions pre-test	0.00	6.84	8.48	0.69
	Number of repetitions post-test	0.00	5.00	4.58	0.28
9	Number of repairs pre-test	0.00	2.43	0.51	0.97
	Number of repairs post-test	0.00	8.59	16.78	0.76

As Table 7.1 shows, initial screening of the score distributions for the nine variables revealed that only three met the criteria of normality and homogeneity. The normality of data is represented by the *p* value (significance) of the Shapiro-Wilkes test, which is greater than 0.05. The homogeneity (equal variance across groups) of the data is indicated by the *p* value (*F*-test) of Levene's test, which is also greater than 0.05. The three variables that met the criteria of normality and homogeneity were the articulation rate, phonation time ratio, and number of pauses. However, only the articulation rate showed a normal distribution and homogeneity in both the pre-test ($p = 0.15$, F-test = 0.69) and post-test ($p = 0.35$, F-test = 0.15). The skewness of the distribution is also lower than 2 with the data seems to skew to the right direction. The phonation time ratio and number of pauses were distributed normally in the pre-test but not in the post-test.

Based on these results, the comparison between the two groups could not be made directly. Therefore, the next step was transforming the data by using some standard transformation tools for right skew data such as square root and logarithm transformations (log10) (Cox, 2007). However, the two transformations could not change the shapes of distributions. As such, the inferential analysis had to be conducted by using non-parametric tests. In addition, to unmask other types of meaningful variation in the data, new statistics based on estimation analysis (Cumming, 2014; Plonsky, 2015) to estimate absolute effects in terms of effect sizes and confidence intervals was used.

7.2 Descriptive Analyses

A descriptive analysis was performed to see the central tendency and spread of the data. The results of the descriptive analysis for the explicit strategy training group (EG) are presented in Table 7.2 (see Table 7.3 for the implicit instruction group (IG).

Table 7.2 Minimum, maximum, mean, and standard deviation of each measure in the explicit group

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Speech rate pre-test	29	1.39	3.60	2.47	0.46
Speech rate post-test	28	1.86	3.37	2.61	0.41
Articulation rate pre-test	29	2.35	5.34	3.55	0.65
Articulation rate post-test	28	2.78	4.58	3.51	0.44
Phonation time ratio pre-test	29	0.42	0.89	0.70	0.10
Phonation time ratio post-test	28	0.61	0.92	0.75	0.07
Mean length of run pre-test	29	2.54	13.43	6.19	1.93
Mean length of run post-test	29	0.00	16.36	6.69	2.67
Number of pauses pre-test	29	0.06	0.38	0.17	0.06
Number of pauses post-test	28	0.06	0.25	0.15	0.05
Number of filled pause pre-test	29	0.00	0.11	0.02	0.03
Number of filled pause post-test	28	0.00	0.03	0.01	0.01
Mean duration of pause pre-test	29	0.25	1.64	0.77	0.26
Mean duration of pause post-test	28	0.43	0.92	0.67	0.12
Repetition pre-test	29	0.00	12.00	1.17	2.35
Repetition post-test	28	0.00	4.00	1.00	1.19
Number of repairs pre-test	29	0.00	0.03	0.01	0.01
Number of repair post-test	28	0.00	0.03	0.01	0.01

It can be seen that there were some differences in the means between the pre-test and post-test for the EG. In terms of the measures of speed, the speech rate, the phonation time ratio, and the length of run, an increased mean score is shown from the pre-test to the post-test. The speech rate's mean increased from *M* = 2.47 in the pre-test to *M* = 2.61 in the post-test. The increase in phonation time ratio was from *M* = 0.70 in the pre-test to *M* = 0.75 in the post-test. Meanwhile, the mean length of run increased from *M* = 6.19 to *M* = 6.69 (pre-test to post-test, respectively). While the standard deviation of the speech rate and phonation time ratio decreased from pre-test to post-test, the mean length of run trended in the opposite direction. The reduced standard deviation in the post-test indicates less variation in the group's performance following instruction. One other measure, the articulation rate, experienced a slight decrease from *M* = 3.55 in the pre-test to *M* = 3.51 in the post-test. The standard deviation of the articulation rate was also lower in the post-test compared to the speech rate and phonation time ratio.

Meanwhile, the dysfluency measures decreased slightly on all four variables (number of pauses, number of filled pauses, mean duration of pauses, and repetition) from the pre-test to the post-test. The biggest decrease among these variables was experienced by the mean duration of pause, dropping from *M* = 0.77 in the pre-test to *M* = 0.67 in the post-test. However, decreases in the four variables were accompanied by increases in the standard deviations, indicating a greater variation within the group in the pre-test.

The results of these descriptive analyses in the EG indicate that the participants' speech fluency experienced an improvement, which is potentially impacted by the explicit strategy training instruction. This potential effect was further analyzed using inferential analysis and estimation analysis (see Chapter 6 Sections 6.6.2 for the details) as discussed in Section 7.3.1 below.

Next, the results of the descriptive analysis on the IG, also involving the same nine variables, are detailed in Table 7.3 and then described.

Table 7.3 Minimum, maximum, mean, and standard deviation of each measure in implicit group

	N	Minimum	Maximum	Mean	Std. Deviation
Speech rate pre-test	25	1.44	4.54	2.83	0.70
Speech rate post-test	25	1.74	4.54	2.84	0.63
Articulation rate pre-test	25	2.87	5.17	3.86	0.68
Articulation rate post-test	25	2.99	5.14	4.02	0.56
Phonation time ratio pre-test	25	0.47	0.88	0.72	0.09
Phonation time ratio post-test	25	0.58	0.88	0.71	0.08
Mean length of run pre-test	25	3.22	21.33	7.22	3.89
Mean length of run post-test	25	3.29	10.93	6.94	2.13
Number of pauses pre-test	25	0.03	0.30	0.16	0.08
Number of pauses post-test	25	0.08	0.30	0.16	0.05
Filled pause pre-test	25	0.00	0.09	0.02	0.02
Filled pause post-test	25	0.00	0.09	0.01	0.02
Mean duration of pause pre-test	25	0.36	1.24	0.67	0.18
Mean duration of pause post-test	25	0.22	1.37	0.71	0.24
Repetition pre-test	25	0.00	10.00	1.68	2.46
Repetition post-test	25	0.00	5.00	1.32	1.46
Number of repairs pre-test	25	0.00	0.04	0.01	0.01
Number of repair post-test	25	0.00	0.06	0.01	0.01

As shown in Table 7.3, the means for phonation time ratio and mean length of run decreased from the pre-test to the post-test. The phonation time ratio's mean dropped from *M* = 0.72 (*SD* = 0.09) to *M* = 0.71 (*SD* = 0.08), and the mean length of run reduced from *M* = 7.22 (*SD* = 3.89) to *M* = 6.94 (*SD* = 2.13). The standard deviations in these two variables also demonstrate that the post-test had a smaller variation than the pre-test, indicating that the post-test data are much closer to the mean than the pre-test data. A slight increase in the mean, followed by a decrease in standard deviation from the pre-test to the post-test is shown for the speech rate (*M* = 2.83, *SD* = 0.70 to *M* = 2.84, *SD* = 0.63, respectively).

Meanwhile, the measures of disruption and dysfluency do not exhibit many changes from the pre-test to the post-test. There was the same mean score (*M* = 0.16) for the number of pauses in both tests and a small difference in the range of standard deviations (*SD* = 0.08 at pre-test and *SD* = 0.05 at post-test). The number of filled pauses was comparable on the

pre-test and the post-test (*M* = 0.01 and *M* = 0.01), while its standard deviation remained the same (*SD* = 0.02). In contrast, the mean and the standard deviation of repetition decreased slightly from the pre-test to the post-test (*M* = 1.68, *SD* = 2.46 to *M* = 1.32, *SD* = 1.46, respectively). Finally, the number of repairs was essentially the same in the pre-test and the post-test (*M* = 0.01, *SD* = 0.01 to *M* = 0.01, SD = 0.01, respectively). On the other hand, the mean of articulation rate increased from the pre-test to the post-test (*M* = 3.86, *SD* = 0.68 to *M* = 4.02, *SD* = 0.56, respectively). An increase in the mean score was also found for the duration of pauses (*M* = 0.67, *SD* = 0.18 in pre-test to *M* = 0.71, *SD* = 0.24 in post-test).

The IG's results indicate that these participants were able to produce a slightly faster speech, as indicated by improvement in the articulation rate. At the same time, they could maintain the same rate of dysfluency and disruption in the speech. While an increase in articulation rate could indicate an improvement, an increase in the mean duration of pauses could also indicate an important stage in participants' language acquisition. Increasing the pause duration could indicate that learners are experimenting with new language and automatizing it during speech production. Therefore, both the increasing articulation rate and the increasing pause duration, are complementary and essential to the learning process. To investigate whether the change or improvement experienced by this group is statistically significant and meaningful, inferential non-parametric tests and estimation analysis were conducted (see Section 7.3.2).

When the results from both groups' descriptive analyses were compared, different trends in performance were found from the pre-test to the post-test. Interestingly, two variables that experienced an increased mean in the EG (phonation time ratio and mean length of run) showed a decrease in the IG. On the other hand, two variables with decreased mean scores in the EG experienced an increase in the IG (articulation rate and mean duration of pauses). Some similarities found in the two groups were in the speech rate and repetition. The speech rate's mean for both groups' post-test increased slightly, when compared to the pre-test, while the repetition experienced a minor decrease. Meanwhile, the mean score of the

number of pauses showed an increase from pre-test to post-test in the EG and remained the same in the IG.

Three variables, the number of filled pauses, repetitions, and repairs, showed almost similar mean scores in the pre-test and post-test. These results were potentially caused by the lower frequencies of these variables in the speech samples, as indicated by the minimum and maximum scores. The minimum scores of the three variables are 0, while the maximum scores range from 3 to 12. These results suggest that these three variables could be omitted from further analysis because they do not contribute to the statistical analysis.

Overall, these descriptive results suggest that both groups may have experienced a potential improvement in their speed fluency (speed rate, articulation rate, phonation time ratio) and breakdown fluency (number of pauses, mean duration of pauses). Despite the different trends shown in the development of each speech fluency measure, both implicit and explicit instruction seems to be comparable in these respects. These findings would be informative for the comparability of both instructional conditions to improve EFL learners' speech fluency in a classroom environment. To investigate the significance and magnitude of differences found in the descriptive analysis, further analysis using a series of non-parametric tests (Mann-Whitney U and Wilcoxon Signed Rank) and the Exploratory Software for Confidence Intervals (ESCI) program (Cumming, 2020) were conducted. The analyses are described in Section 7.3 below.

7.3 Inferential and Estimation Analyses

After descriptive analyses of the data, the inferential and estimation analysis was performed to test whether both groups' performance, as reflected by differences in the mean scores within and between groups, was significant and meaningful. Inferential analysis using non-parametric tests was used to test whether the perceived differences of the mean scores within and between groups were significant. Meanwhile, estimation analysis was applied to reveal the meaningfulness of the differences, which are determined by larger effect sizes and smaller confidence intervals.

Before running both analyses, a visual analysis using bar charts (Figure 7.1) was developed to see the differences between the two groups in the means of the six variables in the pre-test. As the descriptive analyses results above suggested, three of the original nine variables (the number of filled pauses, repetitions, and repairs) were not included. The six variables were displayed in two charts because of the wide range of the scores. This was necessary to clearly visualize the confidence intervals of the mean scores. The first chart (Figure 7.1) contained speech rate, articulation rate, and mean length of run, in which means ranged from 2.47 (EG's mean speech rate) to 7.22 (IG's mean length of run). The second chart (Figure 7.2) also contained phonation time ratio, number of pauses, and mean duration of pauses, with means of less than 1 (see Tables 7.2 and 7.3 for details). The bar charts also confirmed the 95% confidence intervals of the means for each variable.

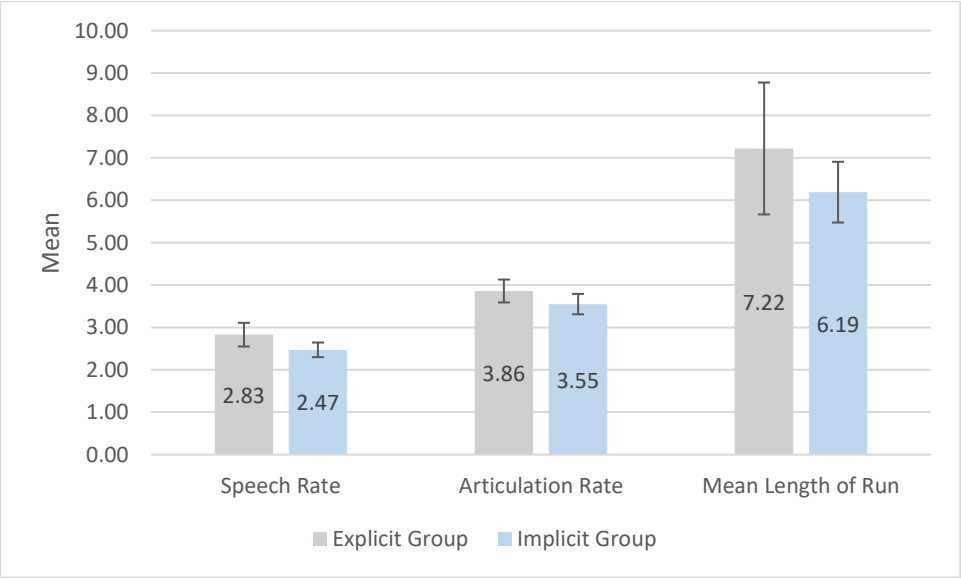


Figure 7.1 Both groups' pre-test means and 95% confidence intervals for speech rate, articulation rate, and mean length of run

Note: 1 signifies the explicit group, and 2 is the implicit group.
Error bars represent 95% confidence intervals of the means.

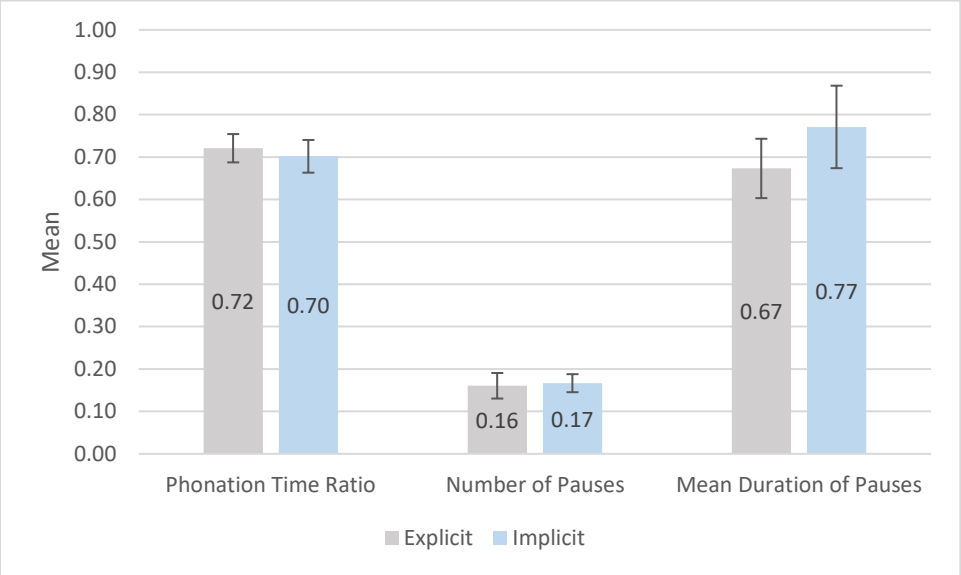


Figure 7.2. Both groups' pre-test means and 95% confidence intervals for phonation time ratio, number of pauses, and mean duration of pauses

Note: 1 signifies the explicit group, and 2 is the implicit group.
Error bars represent 95% confidence intervals of the means.

As seen in Figure 7.1 and Figure 7.2, there are some visible mean differences between groups. The variables in Figure 7.1 (speech rate, articulation rate, and mean length of run) with mean scores above 2.00 show a more visible difference in the charts than those in Figure 7.2 whose mean is below 1.00 (phonation time ratio, number of pauses, and mean duration of pauses). The largest mean difference between groups can be seen in the mean length of run. The IG's mean length of run seems bigger than that of the EG. The IG's mean difference of speech rate is also considerably more than that of the EG. Meanwhile, the smallest mean difference can be seen in the number of pauses; the between group difference is barely visible.

Next, the Mann-Whitney U Test was performed to see whether the differences in mean scores between both groups were significant. As displayed in Table 7.4, the results show that most differences were not statistically significant, except for the speech rate (p = 0.03). However, given its relatively small power (0.57), the significant result produced by speech rate is considered inconclusive. Hence, the Mann-Whitney U Test results indicate that both groups had a comparable level of speech fluency before the treatment.

Table 7.4. The results of the Mann-Whitney U Test in both groups' pre-test scores

Variables	Groups	N	Mean Rank	*Sig.* (*p*)	*Observed Power*
Speech Rate	Implicit	25	32.52	0.03	0.57
	Explicit	29	23.17		
	Total	54			
Articulation Rate	Implicit	25	31.48	0.08	
	Explicit	29	24.07		
	Total	54			
Phonation Time Ratio	Implicit	25	29.26	0.44	
	Explicit	29	25.98		
	Total	54			
Mean Length of Run	Implicit	25	28.98	0.52	
	Explicit	29	26.22		
	Total	54			
Number of Pauses	Implicit	25	25.68	0.43	
	Explicit	29	29.07		
	Total	54			
Mean Duration of pause	Implicit	25	23.68	0.10	
	Explicit	29	30.79		
	Total	54			

Note: Observed power is provided only when the result is significant.

Following the inferential analysis, estimation analysis using ESCI software (Cumming, 2020) was run to estimate the independent mean difference between groups at the time of the pre-test. Table 7.5 details the results of this mean difference estimation, with the effect sizes and 95% confidence intervals for both groups' pre-test scores.

In interpreting effect sizes, Cohen (2013) points out that an effect size of 0.8 or higher is considered large. The data with such effect sizes will be apparent, and the difference will be clearly noticeable even to an untrained eye. An effect size of 0.5 is considered medium, and only a trained eye can notice the difference. Meanwhile, an effect size of 0.2 or smaller will not be apparent in the data without a statistical analysis that has the power to detect it. However, Cohen (2013) also advises that these benchmarks should be used with caution. The causality implied by interpreting an effect size should be assumed only when it can be justified (Coe, 2002). For example, when comparing two groups of students, an effect size's

interpretation should consider their background, including their proficiency level, the treatments they receive, and their demographic situation.

Table 7.5 The results of the independent mean difference estimation of both groups' pre-test

Variables	Group	M	95 % CI		SD	$d_{unbiased}$	N
			Lower	Upper			
Speech Rate	Implicit	2.83	2.55	3.11	0.70		25
	Explicit	2.47	2.30	2.64	0.47		29
	Difference	0.36	0.04	0.68	0.58	0.60	54
Articulation Rate	Implicit	3.86	3.59	4.13	0.68		25
	Explicit	3.55	3.31	3.79	0.65		29
	Difference	0.31	-0.05	0.67	0.66	0.46	54
Phonation Time Ratio	Implicit	0.72	0.69	0.75	0.08		25
	Explicit	0.70	0.66	0.74	0.10		29
	Difference	0.02	-0.03	0.07	0.09	0.20	54
Mean Length of Run	Implicit	7.22	5.66	8.78	3.89		25
	Explicit	6.19	5.47	6.91	1.93		29
	Difference	1.03	-0.62	2.67	3.00	0.34	54
Number of Pauses	Implicit	0.16	0.13	0.19	0.08		25
	Explicit	0.17	0.15	0.19	0.06		29
	Difference	-0.01	-0.04	0.03	0.07	-0.09	54
Mean Duration of Pauses	Implicit	0.67	0.60	0.74	0.18		25
	Explicit	0.77	0.67	0.87	0.26		29
	Difference	-0.10	-0.22	0.03	0.23	-0.43	54

Note: The standardized effect size is d_{unbiased} because the denominator used was SDpooled. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is sometimes called Hedges' g.

As shown in Table 7.5, the pre-test comparison on all variables between groups reveals visible mean differences, with the effect sizes ranging from small ($d_{unbiased}$ = 0.20, M_{diff} = 0.02, 95% CI [-0.03, 0.07] for phonation time ratio) to medium ($d_{unbiased}$ = 0.60, M_{diff} = 0.36, 95% CI [0.04, 0.68] for speech rate). The only variable with a negligible effect size is the number of pauses ($d_{unbiased}$ = -0.09). The mean difference of the number of pauses is M_{diff} = -0.01, 95% CI [-0.04, 0.03].

The mean differences in the comparison are the result of subtracting the mean of the IG from that of the EG; therefore, a negative score occurs when the EG's mean is smaller than that of the IG. Hence, the negative effect sizes are impacted by the negative mean difference. It can be seen from the table that in most comparisons, the IG has higher means than the EG. Mean length of run records the largest difference among the variables. The IG's mean for mean length of run (*M* = 7.22) outscored the EG's mean (*M* = 6.19). However, these results do not automatically suggest that the difference between the two groups is practically meaningful. Further analysis to investigate the magnitude of the difference between means was then conducted by using a scatter plot. The scatter plot was produced by the ESCI to accompany the independent mean difference estimation analysis. As an example, the scatter plot of the mean length of run (Figure 7.2), as a variable with the largest mean difference, was used.

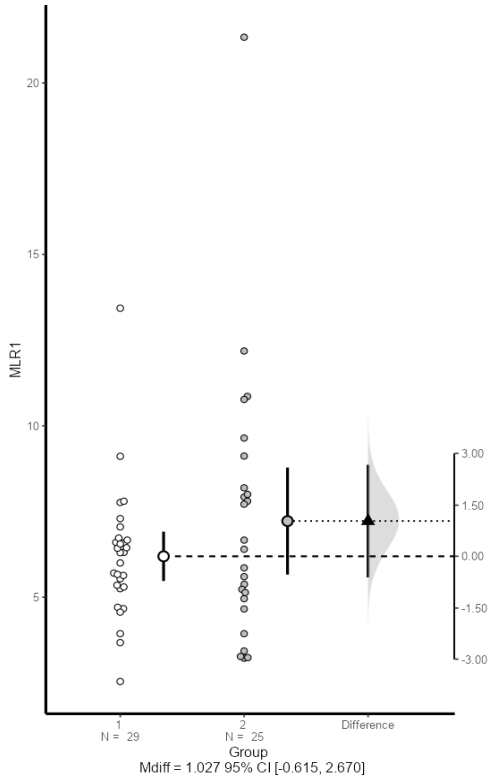


Figure 7.3. Scatter plot of means and 95% confidence intervals (CIs) for mean length of run's pre-test (MLR1) for the two groups

Note: 1. The difference between the group means, with its 95% CI, is shown on a floating difference axis at the right. 2. For the explicit group N = 29, and for the implicit group N = 25.

It can be seen from Figure 7.3 that for mean length of run pre-test (MLR1), the length of the CI on mean difference between the EG and IG is roughly 1.4 times the average of the lengths of the CIs on the two groups' means. This length of the CI is normal for analysis involving two independent groups (Cumming, 2014). However, with these point and interval estimates, the result is considered low for precision because the range of the CI is large, although it could still contribute to a meta-analysis (Cumming, 2014). It can also be seen that the CI of the EG [2.30, 2.68] is much shorter than that of the IG [2.55, 3.11], indicating that the result in the former group is more precise than that of the latter.

Figure 7.3 also shows that the CIs of the two groups overlap for more than half of the average margin of error (MOE), which is represented by one arm of the CI. When two CIs overlap this much, the evidence of a difference between two means is unlikely to be meaningful. Cumming (2014) argues that in order to achieve a meaningful mean difference, the two CIs being compared must not overlap for more than half of the MOE. "If the two groups' CIs just touch or do not overlap, there is reasonable evidence of a population difference" (Cumming, 2014, p. 19). For the rest of the variables, the overlap of the CIs between the two groups is nearly complete, suggesting that the groups were comparable at the time of the pre-test on the variables of interest in the present study (see Appendix 7 for the complete estimation results).

7.3.1. The fluency of learners who receive strategy training in addition to practice

The inferential and estimation analyses were also conducted to compare the pre-test and post-test means of the EG (who received strategy training and practice). The analyses were intended to reveal any possible changes or improvements in learners' speech fluency after the treatment. Before these two analyses, bar charts displaying the means of the EG in the pre-test and post-test were also examined (Figure 7.4 and Figure 7.5). These bar charts show 95% confidence intervals of the mean scores.

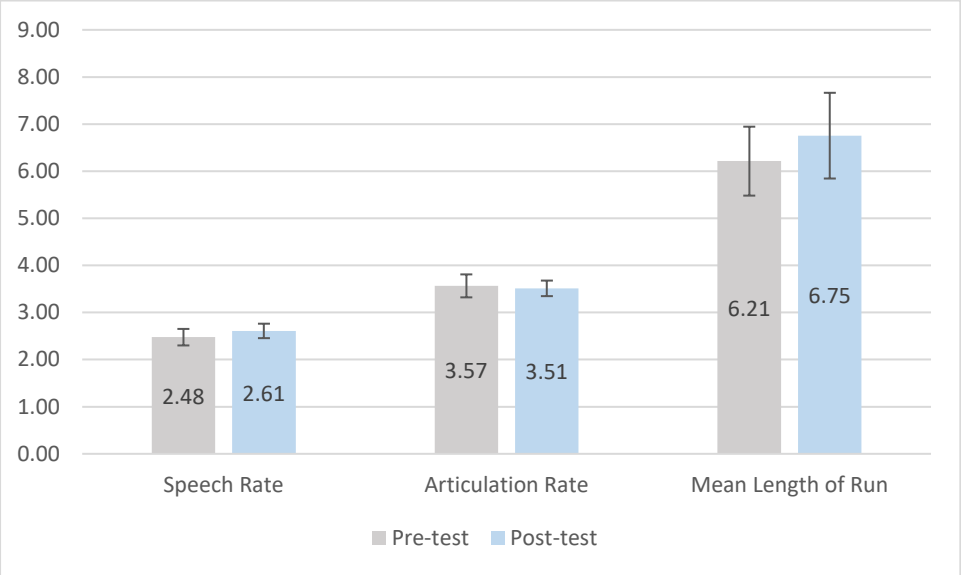


Figure 7.4. The explicit group's pre-test and post-test mean scores for speech rate, articulation rate, and mean length of run

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

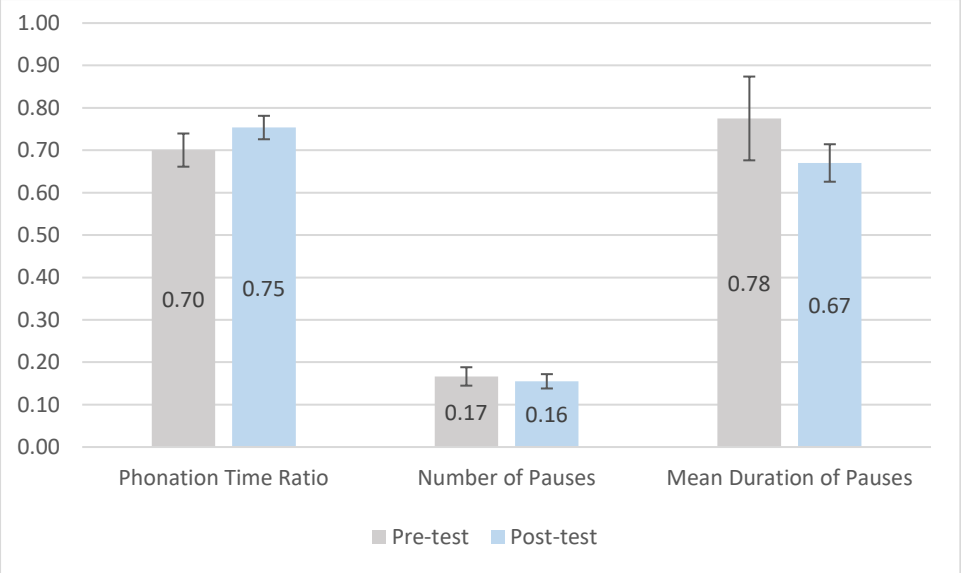


Figure 7.5. The explicit group's pre-test and post-test mean scores for phonation time ratio, number of pauses, and mean duration of pauses

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

Figure 7.4 and Figure 7.5 show that most of the variables in the EG experienced an increase from the pre-test results to the post-test ones, except for the articulation rate. The biggest increase is shown by the mean length of run; this increased from *M* = 6.21 in the pre-test to *M* = 6.75 in the post-test (see Table 7.7). The mean speech rate also shows a visible increase, from M = 2.48 to M = 2.61 (in the pre-test and post-test, respectively). Meanwhile, the increases between the pre-test and the post-test in other variables were much lower (see Table 7.7).

To assess whether the visible differences in the mean between the pre-test and post-test scores were significant, the Wilcoxon Signed Rank Test was performed. As displayed in Table 7.6, most results did not reach a significant difference between the pre-test and post-test scores. However, for one variable (phonation time ratio) a statistically significant difference occurred (z = -2.22, *p* = 0.03) with an observed power of 0.57. Again, this relatively small power suggests that this result in the phonation time ratio is also inconclusive.

Despite these non-significant results, higher mean ranks in the post-tests of the speech rate, phonation time ratio, and mean length of run could indicate that participants in the EG improved in speaking. On the other hand, lower post-test ranks in the number of pauses and mean duration of pauses could also suggest improvement because this indicates that participants produced fewer and shorter pauses in the post-test. Surprisingly, different results were obtained for the articulation rate since the post-test score was lower than that of the pre-test. This condition was caused by a bigger variation in the articulation rate's pre-test data (*Min* = 2.35, *Max* = 5.34, *SD* = 0.65) than its post-test data (*Min* = 2.78, *Max* = 4.58, *SD* = 0.44).

Table 7.6. The results of the Wilcoxon Signed Rank tests for the explicit group

Variables	Tests	N	Mean Rank	*z*	*Sig. (p)*	*Observed Power*
Speech Rate	Pre-test	29	1.61	-1.37	0.17	
	Post-test	28	2.04			
	Total	57				
Articulation Rate	Pre-test	29	2.04	-0.39	0.69	
	Post-test	28	1.93			
	Total	57				

Variables	Tests	N	Mean Rank	*z*	*Sig. (p)*	*Observed Power*
Phonation Time Ratio	Pre-test	29	1.64	-2.22	0.03	0.57
	Post-test	28	2.25			
	Total	57				
Mean Length of Run	Pre-test	29	1.83	-1.48	0.14	
	Post-test	28	2.34			
	Total	57				
Number of Pauses	Pre-test	29	2.29	-1.02	0.31	
	Post-test	28	2.11			
	Total	57				
Mean Duration of Pauses	Pre-test	29	2.18	-1.71	0.08	
	Post-test	28	1.79			
	Total	57				

Note: Observed power is provided only when the result is significant.

To estimate the meaningfulness of the mean difference between the pre-test and post-test scores for the EG, an estimation analysis using ESCI software (Cumming, 2020) was also run. The results are presented in Table 7.7.

Table 7.7 The results of paired mean difference estimation on six fluency measures for the explicit group

Variables	Test	M	95 % CI		SD	$d_{average}$	N
			Lower	Upper			
Speech Rate	Post-test	2.61	2.45	2.77	0.41		28
	Pre-test	2.48	2.29	2.66	0.47		28
	Difference	0.13	-0.06	0.33	0.50	0.29	28
Articulation Rate	Post-test	3.51	3.34	3.69	0.44		28
	Pre-test	3.57	3.31	3.82	0.66		28
	Difference	-0.05	-0.33	0.23	0.72	-0.09	28
Phonation Time Ratio	Post-test	0.75	0.72	0.78	0.07		28
	Pre-test	0.70	0.66	0.74	0.11		28
	Difference	0.05	0.00	0.10	0.13	0.58	28
Mean Length of Run	Post-test	6.75	5.80	7.71	2.45		28
	Pre-test	6.21	5.45	6.98	1.97		28
	Difference	0.54	-0.66	1.74	3.09	0.24	28
Number of Pauses	Post-test	0.16	0.14	0.17	0.05		28
	Pre-test	0.17	0.14	0.19	0.06		28
	Difference	-0.01	-0.04	0.02	0.07	-0.22	28
Mean Duration of	Post-test	0.67	0.62	0.72	0.12		28

Variables	Test	M	95 % CI		SD	$d_{average}$	N
			Lower	Upper			
Pauses	Pre-test	0.78	0.67	0.88	0.27		28
	Difference	-0.11	-0.21	0.00	0.27	-0.5	28

Note: The standardized effect size is $d_{average}$ because the denominator used was SDavg. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is also commonly referred to as Hedges' g. Minus values in $d_{average}$ are caused by higher scores in the pre-test than in the post-test.

As shown in Table 7.7, there are visible increases or changes in the means of most variables in the post-test, with the effect sizes ranging from small ($d_{average}$ = 0.24, M_{diff} = 0.54, 95% CI [-0.66, 1.74] for the mean length of run) to medium ($d_{average}$ = 0.58, M_{diff} = 0.05, 95% CI [0.00, 0.10] for the phonation time ratio). One variable, the articulation rate, experienced a decrease of 0.05, with a negligible effect size ($d_{average}$ =-0.09). Negative scores in the mean differences of the number of pauses and mean duration of pauses indicate an improved performance because the participants produced fewer and shorter pauses in the post-test. The mean difference (improvement) in the number of pauses is M_{diff} = -0.01, with a small effect size ($d_{average}$ = 0.22). For the mean duration of pause, the difference is M_{diff} = -0.11, with a medium effect size ($d_{average}$ = 0.5).

To investigate whether a meaningful difference emerged for the phonation time ratio, as the only variable with a statistically significant result with medium effect sizes, from the pre-test to the post-test, the ESCI's figure which resulted from the estimation analysis was used (Figure 7.6).

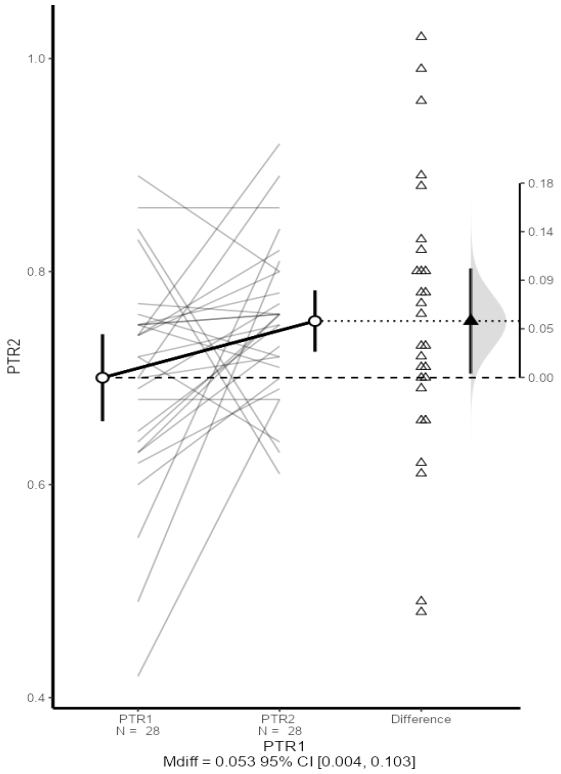


Figure 7.6. Scatter plot of means and 95% confidence intervals (CIs) for phonation time ratio

Note: 1. The difference between the group means, with 95% CI, is shown on a floating difference axis at the right. 2. PTR1 is the pre-test of the phonation time ratio, and PTR2 is the post-test.

Judging from the length of the CIs as shown in Figure 7.6, it seems that the post-test has a more precise result than the pre-test (see the explanation above about a precise result based on the length of the CI in Figure 7.1). The post-test's CI [0.72, 0.78] is shorter than the pre-test's [0.66, 0.74]. However, it can also be seen that the CIs of the two means overlap for more than half of the one arm of the CI (average MOE). This result indicates that the mean difference between the two tests is not too large. For the rest of the variables, the overlap of the CIs between the two scores is nearly complete, suggesting that the magnitude of difference between the pre-test and the post-test of the variables of interest is not meaningful (see Appendix 7 for complete estimation results).

Although in the comparison between the pre-test and post-test means, the difference did not reach statistical significance because of the sample size employed, it is possible to argue that the effect sizes ranging from small to medium could potentially be meaningful for EFL

learners in the study context. This change in effect size (small to medium) could be distinctive for learners' speech fluency development. These learners do not have exposure to the target language environment, so they have little to no opportunity to practice the language outside of the classroom, particularly for the speaking skill. In addition, the learners participating in the present study were students of a third-tier university in Indonesia. Students who enrol in this type of university are usually lower in their academic achievement than those of first and second-tier universities. Therefore, for learners in this cohort, a minor improvement in their speaking skills could be considered meaningful.

When learners improve their speech slightly, their teachers or peers can recognize this easily. For example, the improvement in speech rate, which has a small effect size, could mean that learners could speak more fluently than usual. An improvement in the phonation time ratio and mean duration of pauses with a medium effect size could be considered a substantial shift in learners' speaking development. These improvements would make learners become faster speakers who can speak with fewer and shorter pauses. Therefore, it can be argued that, within the context of the present study, the instruction in strategy training plus the practice had an essential impact on EFL learners' speech fluency (see further discussion regarding the implication of the study in Chapter 9).

7.3.1.1 Proceduralization in the Explicit Group

Three variables were used in combination and analyzed to find evidence of proceduralization. These were the mean length of runs, mean duration of pauses, and phonation time ratio (see Chapter 4 Section 4.3.1). As detailed in Table 7.7, the two variables that showed a meaningful improvement in their mean differences between the pre-test and the post-test are the phonation time ratio ($d_{average}$ = 0.58) and mean duration of pauses ($d_{average}$ = -0.5). Other variable, the mean length of run ($d_{average}$ = 0.24), showed a small difference between its pre-test and post-test scores, indicating a stable development.

The results shown by the three variables could potentially indicate that the participants in the EG show evidence of proceduralization in their language knowledge. This is supported by the fact that the participants could produce a slightly increased stretch of fluent speech

(mean length of run) with less time for pausing (improved phonation time ratio and mean duration of pause). This result is in line with de Jong and Perfetti's (2011) conclusion from their study regarding proceduralization (see Chapter 4 Section 4.3.1).

7.3.1.2 Raters' Judgement for oral proficiency

Raters' judgements were used to provide an independent measure of gains in participants' oral proficiency. Two raters (see Section 6.5.4.3) judged participants' performances in the pre-test and post-test based on Brown's (2001) and Choi's (2005) oral proficiency rating scales. Brown's holistic rating covers only the general proficiency scale. Choi's analytic rating includes five independent scales: pronunciation, discourse, vocabulary, grammar, and complexity (Section 6.4 and Appendix 5 for more information). The results from both raters were then compared to ascertain if there was any meaningful improvement in participants' oral performance when comparing their pre-test to their post-test scores.

Before the comparison, the internal reliability of the scores from each rater was determined using Cronbach's alpha to ensure the validity of the ratings. Rater 1's scores yielded a Cronbach's alpha value of 0.94. For Rater 2, the value was 0.90, and the combined scores of both raters also resulted in a Cronbach's alpha value of 0.90. This result suggests that the obtained scores were internally reliable because the values exceed 0.70 as the cut-off for a suitable Cronbach's alpha value (Field, 2017). However, the scores did not reach an acceptable value for inter-rater agreement or for the degree of agreement between the two raters. The Cohen's Kappa value of 0.02, which was achieved after collapsing categories, was still far from indicating an acceptable inter-rater agreement. As suggested by Cohen (1960), this value represents only a slight agreement. The lack of agreement between the two raters happened in all almost categories across the two rubrics. Raters' unreliability could be caused by differences in their educational background, or bias due to familiarity with participants (Batty & Stewart, 2016).

Despite this lack of agreement between raters, the raters' judgements were analyzed to investigate any improvements (gains) in connection with the explicit strategy training instruction. For analysis purposes, the scales from both raters were combined and divided

by two (number of raters) to obtain the average mean scores. These mean scores were then analyzed by using a bar chart for visual analysis, the Wilcoxon Signed Rank Test, for the test of significance, and the paired mean difference estimation of ESCI to determine meaningful differences between the pre-test and the post-test scores of the EG.

First, a bar chart accompanied by the error bars which represent the 95% confidence intervals of the scales was created (Figure 7.7).

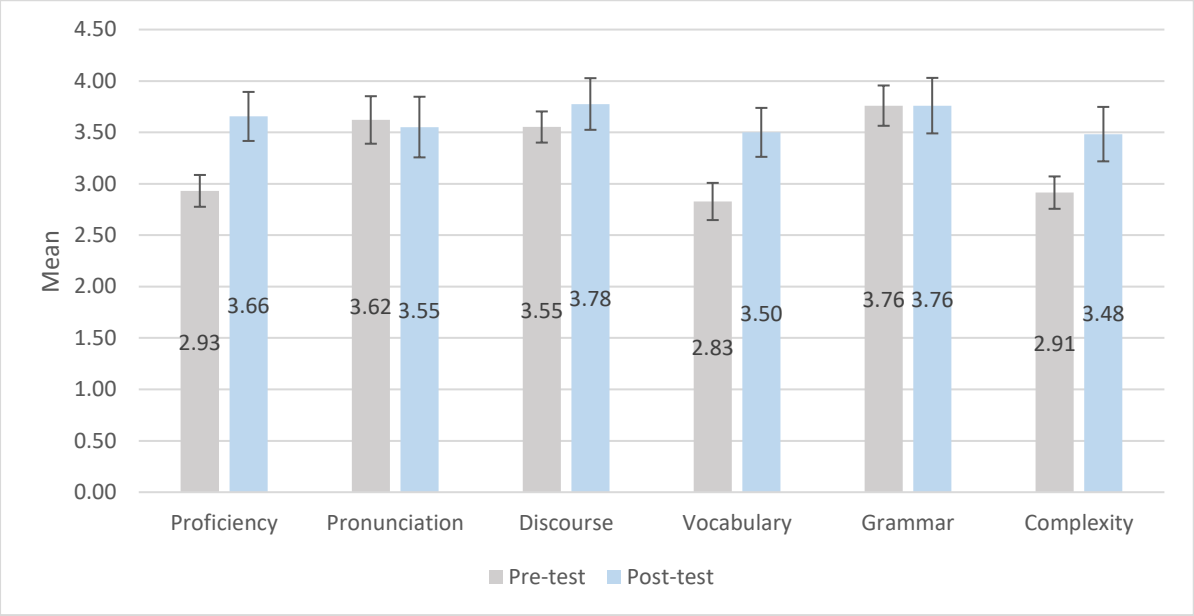


Figure 7.7. The explicit group's pre-test and post-test mean scores for the oral proficiency scales

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

As can be seen in Figure 7.7, for most of the scales, there is a visible increase in the mean from the pre-test to the post-test, except for pronunciation. The general proficiency scale appears to have the biggest increase, followed by vocabulary and complexity. To investigate whether the visible increases in the mean scores of the oral proficiency scales are statistically significant, the Wilcoxon Signed Rank Test was conducted, and the results are presented in Table 7.8.

Table 7.8. The results of the Wilcoxon Signed Rank Test of the explicit group's oral proficiency scales

Variables	Tests	N	Mean Rank	*z*	*p*	*Observed Power*
Proficiency	Pre-test	29	1.14	-3.73	< 0.01	1.00
	Post-test	28	1.86			
	Total	57				
Pronunciation	Pre-test	29	1.52	-0.46	0.65	
	Post-test	28	1.48			
	Total	57				
Discourse	Pre-test	29	1.29	-2.08	0.04	0.55
	Post-test	28	1.71			
	Total	57				
Vocabulary	Pre-test	29	1.12	-3.77	< 0.001	1.00
	Post-test	28	1.88			
	Total	57				
Grammar	Pre-test	29	1.41	-0.49	0.62	
	Post-test	28	1.59			
	Total	57				
Complexity	Pre-test	29	1.10	-3.20	< 0.01	0.97
	Post-test	28	1.90			
	Total	57				

Note: Observed power is provided only when the result is significant.

It can be seen from Table 7.8 that four scales in the list resulted in a statistically significant difference. These scales were proficiency (z = -3.73, p < 0.01), discourse (z = -2.08, p = 0.04), vocabulary (z = -3.77, p < 0.01), and complexity (z = -3.20, p < 0.01). The average observed power of these scales was acceptable (> 0.80), except for discourse, whose observed power was only 0.55. The differences in two other scales, pronunciation and grammar, did not reach statistical significance. These results support the visual analysis results of the bar chart (Figure 7.7) and indicate that the EG's oral proficiency improved as the effect of the strategy training instruction.

Following the test of significance, a paired mean difference estimation analysis was conducted on the oral fluency scales to determine the magnitudes of the effects (effect sizes). The estimation analysis is detailed in Table 7.9.

Table 7.9 The results of paired mean difference estimation of oral fluency scales for the explicit group

Variable	Condition	M	95 % CI		SD	$d_{average}$	N
			Lower	Upper			
Proficiency	Post-test	3.66	3.41	3.90	0.64		29
	Pre-test	2.93	2.77	3.09	0.42		29
	Difference	0.72	0.46	0.99	0.69	1.32	29
Pronunciation	Post-test	3.55	3.25	3.85	0.79		29
	Pre-test	3.62	3.38	3.86	0.62		29
	Difference	-0.07	-0.32	0.18	0.66	0.1	29
Discourse	Post-test	3.78	3.52	4.03	0.68		29
	Pre-test	3.55	3.40	3.71	0.41		29
	Difference	0.22	0.01	0.44	0.56	0.4	29
Vocabulary	Post-test	3.50	3.26	3.74	0.64		29
	Pre-test	2.83	2.64	3.01	0.49		29
	Difference	0.67	0.43	0.92	0.65	1.17	29
Grammar	Post-test	3.76	3.48	4.04	0.73		29
	Pre-test	3.76	3.56	3.96	0.53		29
	Difference	0.00	-0.27	0.27	0.71	0	29
Complexity	Post-test	3.48	3.21	3.75	0.71		29
	Pre-test	2.91	2.75	3.08	0.42		29
	Difference	0.57	0.28	0.86	0.75	0.96	29

Note: The standardized effect size is $d_{average}$ because the denominator used was SDavg. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is also commonly referred to as Hedges' g. Minus values in $d_{average}$ are caused by higher scores in the pre-test than in the post-test.

It can be seen from Table 7.8 that four scales in the list had a meaningful difference between the pre-test and the post-test. The effect sizes average of those scales ranges from 0.4 (near medium, for discourse) to 1.32 (large, for proficiency). The general proficiency scale appears to have the biggest mean difference (M_{dif} = 0.72 [0.48, 0.99]) and effect size (d_{avg} = 1.32). In the pre-test, participants' mean score on the general proficiency scale was *M* = 2.93. This score is within Level 2 of the rubric (see Appendix 5 for information regarding the rubric), and the speaking skill at this level is called developing speaking (Brown, 2001). In the post-test, the general proficiency's mean improved to Level 3 (*M* = 3.66), which is categorized as competent speaking. The second scale, representing a very large effect size (d_{avg} = 1.17), is vocabulary (M_{dif} = 0.67 [0.43, 0.92]). In this aspect, participants' speaking skills improved from Level 2 (*M* = 2.83) in the pre-test to Level 3 (*M* = 3.50). Level 2 in this

analytic rubric means that participants' vocabulary is less varied, with many words used repeatedly. Level 3 indicates varied vocabulary with some use of idiomatic expressions. The other scale which also resulted in a large effect size is complexity (d_{avg} = 0.96, M_{dif} = 0.57 [0.28, 0.86]). The speaking skill in this scale also improved from Level 2 (M = 2.91) to Level 3 (M = 3.48).

Meanwhile, for the other two scales, pronunciation and grammar, a different trend is shown since their scores did not increase between the pre-test and the post-test. In fact, pronunciation scores decreased. Apart from these two scores, it can be concluded that based on raters' judgements, the participants from this group improved somewhat in their oral proficiency.

7.3.2 The fluency of learners who received only practice

The same series of analyses was also applied to the IG to determine the effect of implicit instruction on learners' speech fluency. This comprised visual analysis using bar charts, the test of significant difference, and estimation analysis. All analyses were conducted by comparing the pre-test and post-test mean scores of the six speech fluency measures: speech rate, articulation rate, phonation time ratio, mean length of run, number of pauses, and mean duration of pauses.

Before performing the test of significant difference and estimation analyses, a visual analysis using a bar chart was done. Figure 7.8 and Figure 7.9 show the bar charts for the pre-test and post-test mean scores of the six variables being investigated.

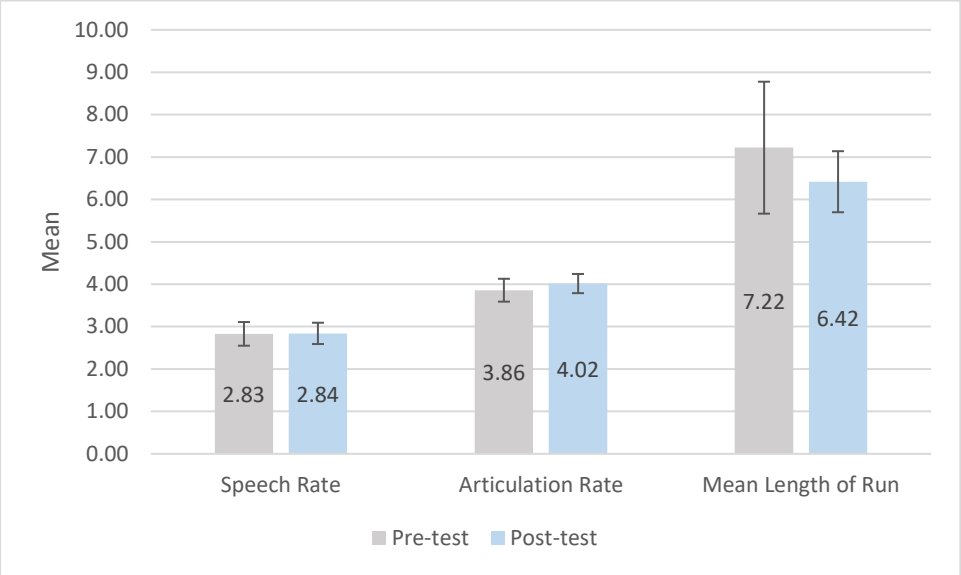


Figure 7.8. The implicit group's pre-test and post-test mean scores for speech rate, articulation rate, and mean length of run

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

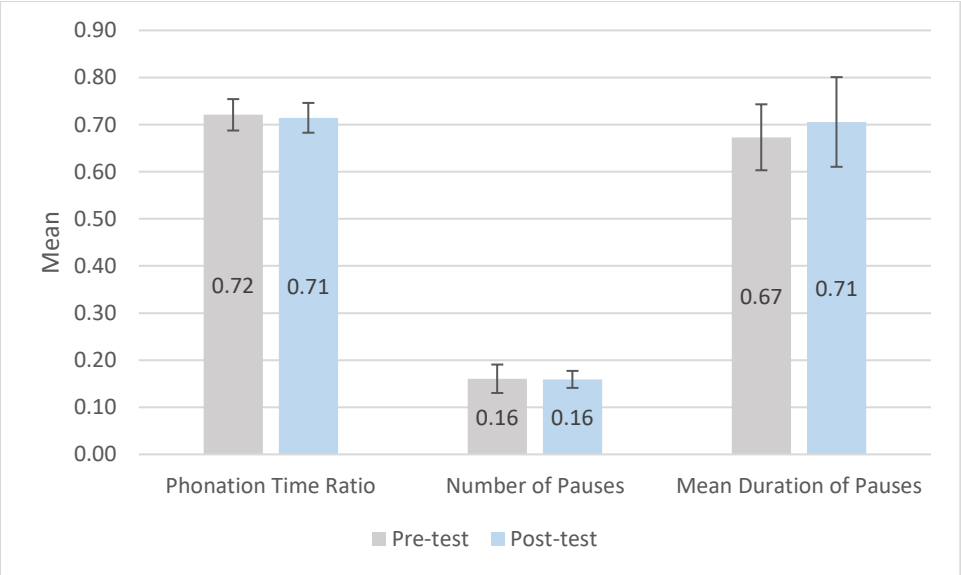


Figure 7.9. The implicit group's pre-test and post-test mean scores for phonation time ratio, number of pauses, and mean duration of pauses

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

The bar charts in Figure 7.8 and Figure 7.9 show that the means of the speech fluency measures do not change much from the pre-test to the post-test. A minor increase can be seen in the mean articulation rate and mean duration of pauses. However, the increase shown by the mean duration of pauses indicates an increase in the pausing time, which does not correlate with an improvement of speech fluency. In addition, a visible decrease can also be seen in the mean length of run, indicating that participants' fluent speech between pauses is shorter in the post-test than in the pre-test.

Further analysis to investigate the significance of the mean difference between the pre-test and the post-test of the IG was then conducted using the Wilcoxon Signed Rank Test. The results are presented in Table 7.10.

Table 7.10. The results of Wilcoxon Signed Rank Test of the speech fluency measures in the implicit group

Variables	Tests	N	Mean Rank	*z*	*P*
Speech Rate	Pre-test	25	12.60	-0.34	0.74
	Post-test	25	13.64		
	Total	50			
Articulation Rate	Pre-test	25	11.40	-1.31	0.19
	Post-test	25	14.07		
	Total	50			
Phonation Time Ratio	Pre-test	25	13.08	-0.2	0.84
	Post-test	25	11.92		
	Total	50			
Mean Length of Run	Pre-test	25	14.69	-0.77	0.44
	Post-test	25	11.17		
	Total	50			
Number of Pauses	Pre-test	25	13.50	-0.01	0.99
	Post-test	25	12.54		
	Total	50			
Mean Duration of Pauses	Pre-test	25	12.73	-0.77	0.44
	Post-test	25	13.40		
	Total	50			

Table 7.10 shows that in the IG, for no variables was there a statistically significant difference. The IG had an improved mean speech rate and mean articulation rate. These two results could indicate improvement in participants' rates of speech. On the contrary, the mean phonation time ratio and mean length of run showed some decreases for the post-test scores compared with those of the pre-test. These results indicate that participants in this group paused more and with a longer duration in the post-test. This was supported by the increased mean number of pauses and mean duration of pauses.

Next, estimation analysis was performed using the ESCI software to investigate the magnitudes of the difference (effect sizes) in mean scores between the IG's pre-test and post-test. The results are detailed in Table 7.11.

Table 7.11 The results of paired mean difference estimation on six fluency measures for the implicit group

Variables	Group	M	95 % CI		SD	$d_{average}$	N
			Lower	Upper			
Speech Rate	Post-test	2.84	2.58	3.10	0.63		25
	Pre-test	2.83	2.54	3.12	0.70		25
	Difference	0.01	-0.36	0.38	0.90	0.02	25
Articulation Rate	Post-test	4.02	3.78	4.25	0.57		25
	Pre-test	3.86	3.58	4.14	0.68		25
	Difference	0.16	-0.19	0.51	0.85	0.25	25
Phonation Time Ratio	Post-test	0.71	0.68	0.75	0.08		25
	Pre-test	0.72	0.69	0.76	0.08		25
	Difference	-0.01	-0.05	0.04	0.11	-0.08	25
Mean Length of Run	Post-test	6.42	5.67	7.16	1.80		25
	Pre-test	7.22	5.62	8.83	3.89		25
	Difference	-0.80	-2.46	0.85	4.01	-0.26	25
Number of Pauses	Post-test	0.16	0.14	0.18	0.05		25
	Pre-test	0.16	0.13	0.19	0.08		25
	Difference	0.00	-0.04	0.03	0.08	-0.02	25
Mean Duration of Pauses	Post-test	0.71	0.61	0.80	0.24		25
	Pre-test	0.67	0.60	0.75	0.18		25
	Difference	0.03	-0.06	0.13	0.23	0.15	25

Note: The standardized effect size is $d_{average}$ because the denominator used was SDavg. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is also commonly referred to as Hedges' g. Minus values in $d_{average}$ are caused by higher scores in the pre-test than in the post-test.

Table 7.11 shows that the overall effect sizes of the mean differences are small. Two variables with small effect sizes are the articulation rate ($d_{average}$ = 0.25) and the mean length of run ($d_{average}$ = -0.26). The other four variables have negligible effect sizes ranging from $d_{average}$ = -0.02 (number of pauses) to 0.15 (mean duration of pauses). Mean duration of pauses has an increased mean from the pre-test (*M* = 0.67) to the post-test (*M* = 0.71). Similarly, the speech rate also had a slightly increased mean from *M* = 83 (pre-test) to *M* = 84 (post-test). The mean number of pauses in both tests remains the same (*M* = 0.16). On the contrary, the phonation time ratio, which has a negligible effect size ($d_{average}$ = -0.08), shows a slight decrease in scores from the pre-test (M = 0.72) to the post-test (M = 0.71). These results suggest that participants in this group paused more and with longer durations in the post-test.

In general, the results for the implicit instruction group may indicate a minor improvement in fluency from the pre-test to the post-test, despite these being non-significant results due to the sample size employed. This minor improvement could be caused by this group's slightly higher speech fluency than the EG before the treatments began (see Table 7.4 and Section 7.3). This could also be due to the impact of task repetition with similar (but not the same) tasks (see Chapter 6 Section 6.4) in implicit instruction for EFL learners in the study. Further discussion regarding the impact of this implicit instruction is detailed in Chapter 9 Section 9.1.

Again, a minor improvement might be considered essential for these learners because of their limitation as third-tier university students in an EFL context. They almost never use English in their daily routines, except in the English classroom. They do not have exposure to the target language environment, such as in an ESL context, which would have provided an opportunity to practise. Therefore, any improvement in their language skills, especially speaking, would be quite important for their language learning development.

7.3.2.1 Proceduralization in the Implicit Group

Three variables are indicative of proceduralization: the mean length of run, mean duration of pause, and phonation time ratio. As detailed in Table 7.10 and Table 7.11, no significant

difference was evident between the pre-test and post-test in these three variables and no meaningful effect sizes were detected. In fact, for these variables a decrease in the post-tests means occurred. Therefore, it could be concluded that there was no indication of proceduralization shown by the participants in the IG.

7.3.2.2 Raters' judgment for oral proficiency

The same two raters also judged participants in the IG based on the same oral proficiency rating scales (see Section 7.3.1.2 regarding raters' judgement in the EG). Similar to the analysis performed for the EG, the analysis of oral fluency scales for the IG also involved bar charts, a test of significant difference, and estimation analysis. The bar chart was used to visualize the difference between the pre-test and post-test means for the IG and is presented in Figure 7.10.

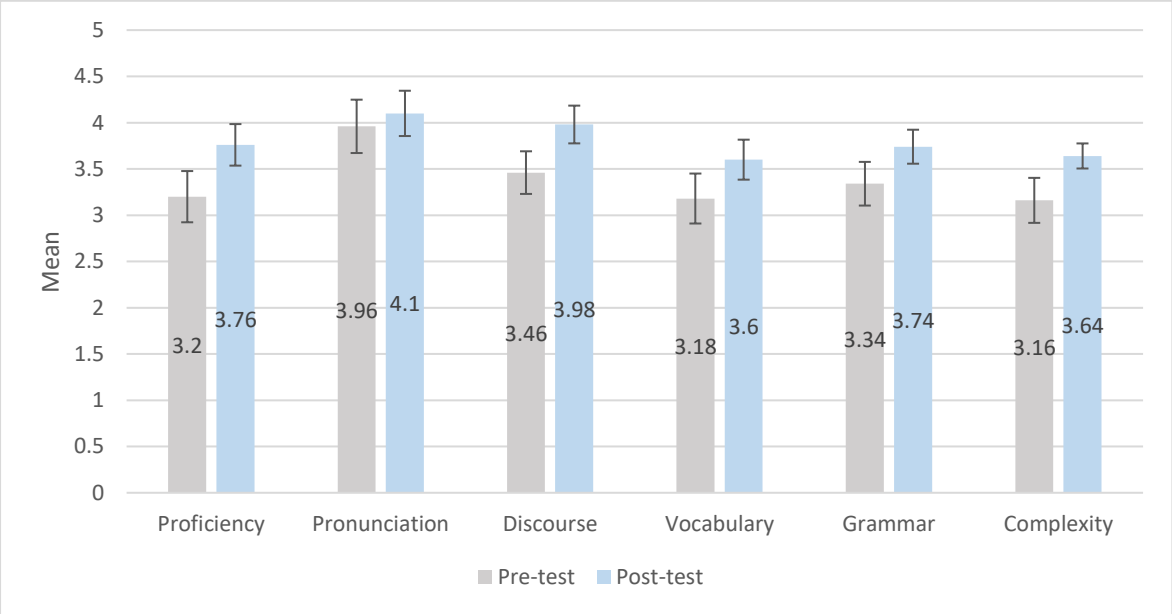


Figure 7.10. The implicit group's pre-test and post-test mean scores for the oral proficiency scales

Note: 1 denotes the pre-test, and 2 is the post-test.
Error bars represent 95% confidence intervals of the means.

For most of the oral proficiency scales in Figure 7.7 an increase in the mean from the pre-test to the post-test is visible, with general proficiency and discourse seeming to have the highest. A visible increase is also evident for complexity while in other three scales

(pronunciation, vocabulary, and grammar), the increase seems to be moderate. Next, the Wilcoxon Signed Rank Test was conducted to investigate the significance of this mean difference (increase) between the pre-test and post-test scores for this IG (see Table 7.12).

Table 7.12. The results of Wilcoxon Signed Rank Test for the implicit group's oral proficiency scales

Variables	Tests	N	Mean Rank	*z*	*Sig. (p)*	*Observed Power*
Proficiency	Pre-test	25	1.16	-3.67	< 0.01	0.99
	Post-test	25	1.84			
	Total	50				
Pronunciation	Pre-test	25	1.48	-1.07	0.28	
	Post-test	25	1.52			
	Total	50				
Discourse	Pre-test	25	1.20	-3.49	< 0.01	0.99
	Post-test	25	1.80			
	Total	50				
Vocabulary	Pre-test	25	1.26	-3.08	< 0.01	0.94
	Post-test	25	1.74			
	Total	50				
Grammar	Pre-test	25	1.22	-3.35	< 0.01	0.97
	Post-test	25	1.78			
	Total	50				
Complexity	Pre-test	25	1.20	-3.41	< 0.01	0.99
	Post-test	25	1.80			
	Total	50				

Note: Observed power is provided only when the result is significant.

Figure 7.12 reveals that for five variables, a statistically significant difference in their means exists between the pre-test and post-test. The level of significance (*p*) of these five is less than 0.01. The observed power is also larger than 0.80. The only variable for which there is not any significant difference is pronunciation ($z = -1.07$, $p = 0.28$). This indicates that the implicit instruction improved the participants' oral proficiency, according to the raters.

To determine the magnitudes of the effects (effect sizes), a paired mean difference estimation analysis using ESCI was conducted on the oral fluency scales for this IG (see Table 7.12).

Table 7.13 The results of paired mean difference estimation of oral fluency scales for the implicit group

Variable	Condition	M	95 % CI		SD	$d_{average}$	N
			Lower	Upper			
Proficiency	Post-test	3.76	3.53	3.99	0.56		25
	Pre-test	3.20	2.91	3.49	0.69		25
	Difference	0.56	0.34	0.78	0.53	0.87	25
Pronunciation	Post-test	4.10	3.85	4.35	0.61		25
	Pre-test	3.96	3.66	4.26	0.72		25
	Difference	0.14	-0.12	0.40	0.62	0.21	25
Discourse	Post-test	3.98	3.77	4.19	0.51		25
	Pre-test	3.46	3.22	3.70	0.58		25
	Difference	0.52	0.30	0.74	0.53	0.94	25
Vocabulary	Post-test	3.60	3.38	3.82	0.54		25
	Pre-test	3.18	2.90	3.46	0.68		25
	Difference	0.42	0.18	0.66	0.57	0.68	25
Grammar	Post-test	3.74	3.55	3.93	0.46		25
	Pre-test	3.34	3.10	3.58	0.59		25
	Difference	0.40	0.19	0.61	0.50	0.74	25
Complexity	Post-test	3.64	3.50	3.78	0.34		25
	Pre-test	3.16	2.91	3.41	0.61		25
	Difference	0.48	0.25	0.71	0.55	0.96	25

Note: The standardized effect size is $d_{average}$ because the denominator used was SDavg. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is sometimes also called Hedges' g. Minus values in $d_{average}$ is caused by higher scores in the pre-test than the post-test.

Table 7.13 shows that for five scales, a potentially meaningful difference between the pre-test and the post-test exists. The effect sizes range from 0.68 (medium, for vocabulary) to 0.96 (large, for complexity). For participants in this IG, the complexity scale has the largest effect size ($d_{average}$ = 0.96) and the second biggest mean difference (M_{dif} = 0.48 [0.25, 0.71]). Improvement in participants' speaking skills in this aspect is still within Level 3 (M = 3.16 in the pre-test to M = 3.64 in post-test).

Similarly to the EG, the general proficiency scale for the IG manifested the largest mean difference between the pre-test and post-test (M_{dif} = 0.56 [0.34, 0.78]) with a large effect size ($d_{average}$ = 0.87). In the pre-test, participants' mean score on the general proficiency scale was M = 3.20 (Level 3). In the post-test, the general proficiency's mean improved to M

= 3.76 (Level 3). Meanwhile, pronunciation was the lowest among all of the scales in terms of the effect size ($d_{average}$ = 0.21) and mean difference (M_{dif} = 0.14 [-0.12, 0.40]) between the pre-test (M = 3.96) and post-test (M = 4.10). Therefore, it can be concluded that, based on the raters' judgement, there was a meaningful improvement in IG participants oral proficiency.

Although this did not result in a shift from one level to another (except for pronunciation) the mean difference between the pre-test and post-test scores for the IG members was comparable to those in the EG. For EFL learners in the context of the current study, the extent of the mean differences and effect sizes reflects a major shift in their oral proficiency development.

7.3.3 Comparing the explicit strategy training and implicit instruction

A series of analyses was also performed to compare the impact of both instructional conditions on learners' speech fluency and oral proficiency. This consisted of a visual analysis using bar charts, an inferential analysis using non-parametric tests, and estimation analysis with ESCI. For this purpose, each group's gain in all six speech fluency measures was compared. The gain was taken to be the mean difference between the pre-test and the post-test scores. The post-test score of each speech fluency measure was subtracted by its pre-test score to achieve the gain.

First, a bar chart that reflects both groups' speech fluency gains was created. Similar to the analysis on the pre-test data comparison (see Section 7.3), two charts were used to display the six variables because of the wide range of the scores. The first chart (Figure 7.11) describes the data for the speech rate, articulation rate, and mean length of run. The second chart (Figure 7.12) describes the data for the phonation time ratio, number of pauses, and mean duration of pauses. The bar charts also give the 95% confidence intervals of the mean scores in each variable.

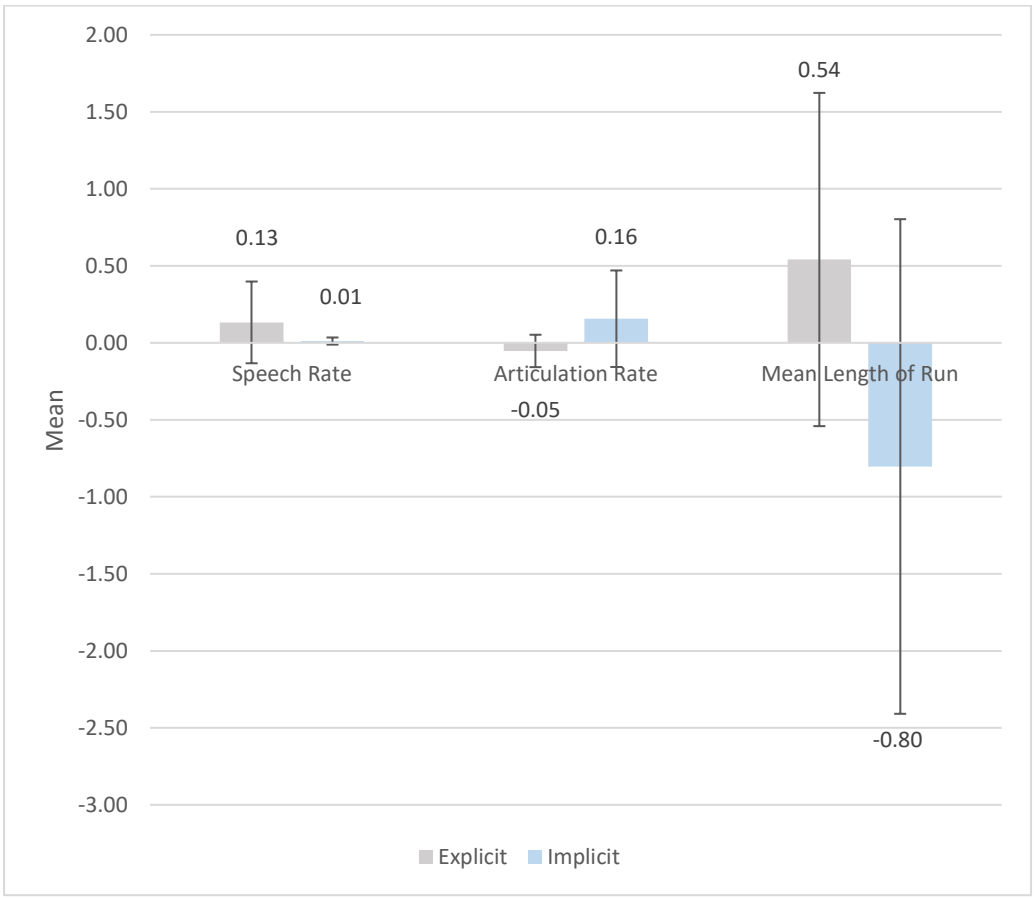


Figure 7.11. Both groups' gains (means) and 95% confidence intervals of speech rate, articulation rate, and mean length of run

Note: 1 denotes the explicit group, and 2 is the implicit group.
Error bars represent 95% confidence intervals of the means.

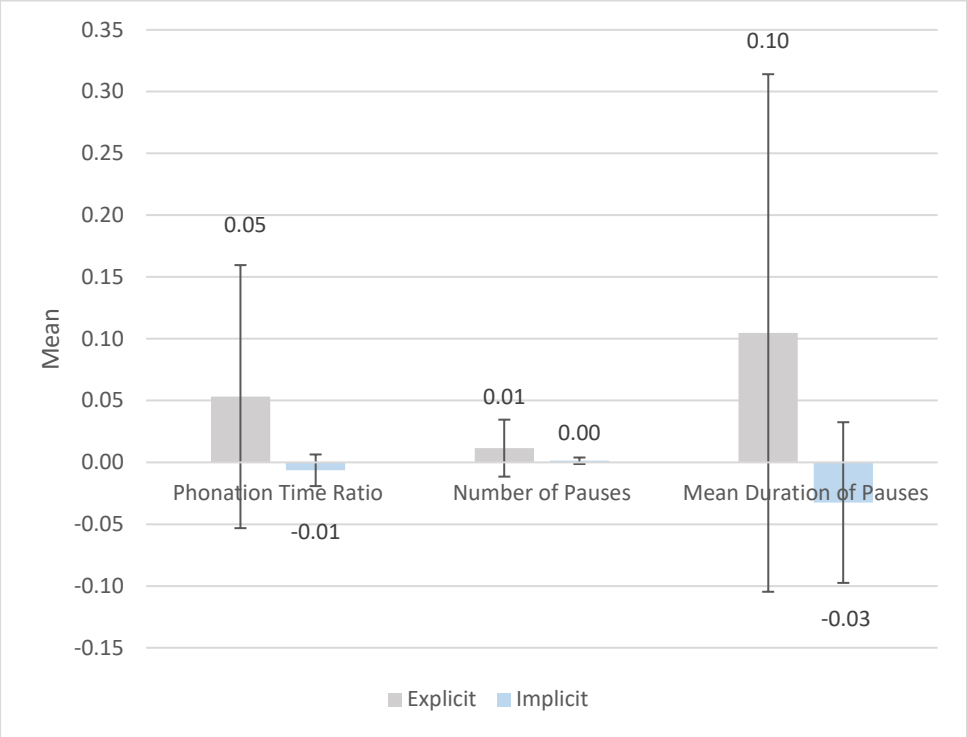


Figure 7.12. Both groups' gains (means) and 95% confidence intervals of phonation time ratio, number of pauses, and mean duration of pauses

Note: 1 denotes the explicit group, and 2 is the implicit group.
Error bars represent 95% confidence intervals of the means.

As can be seen from Figure 7.11 and Figure 7.12, both groups have a different gain in all six variables. The error bars are very large due to the standard deviations, which are much bigger than the means. These bigger standard deviations are the result of how the gain variables were created: by subtracting the post-tests means with the pre-tests'. The new gain variables appear to have many outliers and low values due to the small gains achieved by each group. For example, the number of pauses, speech rate, and articulation rate have means close to zero.

Figure 7.11 shows that the EG's gain is bigger than the IG's in their mean speech rate and mean length of run. The EG also has bigger gains than the IG in their mean phonation time ratio, mean number of pauses, and mean duration of pauses (Figure 7.12). On the contrary, the IG gains seem to be bigger than the EG in their mean articulation rate (Figure 7.11). The EG has a gain that is lower than zero in their articulation rate (Figure 7.11), indicating that

for this variable a decline occurred from the pre-test to the post-test. Meanwhile, the IG also seems to have experienced a decline in terms of three variables: the mean length of run (Figure 7.11), mean phonation time ratio, and mean duration of pauses (Figure 7.12). In general, the results of the bar charts analysis indicate that each form of instruction seems to have a different impact on learners' speech fluency.

Next, the differences in the gain in speech fluency measures as indicated by the bar charts were further analyzed using the Mann-Whitney U Test to investigate any possible significant differences between the groups (see Table 7.14).

Table 7.14. The results of the Mann-Whitney U Test of speech fluency gain for both groups

Variables	Groups	N	Mean Rank	*z*	Sig. (*p*)
Speech Rate	Explicit	28	27.07	-0.036	0.97
	Implicit	25	26.92		
	Total	53			
Articulation Rate	Explicit	28	24.39	-1.301	0.19
	Implicit	25	29.92		
	Total	53			
Phonation Time Ratio	Explicit	28	30.32	-1.658	0.10
	Implicit	25	23.28		
	Total	53			
Mean Length of Run	Explicit	29	28.79	-0.891	0.37
	Implicit	25	25.00		
	Total	54			
Number of Pauses	Explicit	28	28.14	-0.570	0.57
	Implicit	25	25.72		
	Total	53			
Mean Duration of Pause	Explicit	28	30.36	-1.675	0.09
	Implicit	25	23.24		
	Total	53			

As reported in Table 7.14, the differences between the two instructional groups in speech fluency gains did not reach any statistical significance. This indicates that the effect of both instructional conditions on speech fluency is comparable for the EFL learners in the context of the study.

Despite the non-significant results generated by the Mann-Whitney U Test, an estimation analysis was still performed to analyse the magnitude of the difference. The results of the independence mean difference analysis for the speech fluency measures is presented in Table 7.15.

Table 7.15 The results of independent mean difference analysis on speech fluency gains, comparison between the explicit group and the implicit group

Measure	Group	M	95 % CI		SD	$d_{unbiased}$	N
			Lower	Upper			
Speech rate	Implicit	0.01	-0.35	0.37	0.90		25
	Explicit	0.13	-0.06	0.32	0.50		28
	Difference	-0.12	-0.52	0.28	0.72	-0.17	53
Articulation Rate	Implicit	0.16	-0.18	0.50	0.85		25
	Explicit	-0.05	-0.33	0.22	0.72		28
	Difference	0.21	-0.22	0.64	0.78	0.26	53
Phonation Time Ratio	Implicit	-0.01	-0.05	0.04	0.11		25
	Explicit	0.05	0.00	0.10	0.13		28
	Difference	-0.06	-0.13	0.01	0.12	-0.49	53
Mean Length of Run	Implicit	-0.80	-2.41	0.81	4.01		25
	Explicit	0.54	-0.63	1.71	3.09		28
	Difference	-1.34	-3.31	0.62	3.55	-0.37	53
Number of Pauses	Implicit	0.00	-0.03	0.04	0.08		25
	Explicit	0.01	-0.02	0.04	0.07		28
	Difference	-0.01	-0.05	0.03	0.08	-0.13	53
Mean Duration of Pauses	Implicit	-0.03	-0.13	0.06	0.23		25
	Explicit	0.10	0.00	0.21	0.27		28
	Difference	-0.14	-0.28	0.00	0.25	-0.53	53

Note: 1. The standardized effect size is $d_{unbiased}$ because the denominator used was SDpooled. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is sometimes also called Hedges' g. 2. Negative mean scores in a group indicate that the scores in the post-test are lower than the pre-test.

Table 7.15 shows that the mean differences of the gains are varied between the two groups, with the effect sizes ranging from small to medium. A negative score in the mean difference and effect sizes indicates that the gain in the EG is bigger than in the IG. In most of the estimations, the EG has higher mean gains than the IG. The EG has a negative value, indicating a decreased mean, in the articulation rate (M_{gain} = -0.05), while the IG's mean decreased in the phonation time ratio (M_{gain} = -0.01), mean length of run (M_{gain} = -0.80), and

mean duration of pause (M_{gain} = -0.03). The effect sizes of the mean differences in these four variables range from small to medium. The largest effect size ($d_{unbiased}$ = -0.53) is found for the mean duration of pause, with a mean difference of -0.14, indicating a larger gain for the EG. The phonation time ratio also indicates the same condition, in which the EG outscores the IG (M_{diff} = -0.06, [0.00, 0.10], $d_{unbiased}$ = -0.49). Meanwhile, for the number of pauses, there is the lowest mean difference (M_{dif} = -0.01, [-0.05, 0.03) and a negligible effect size ($d_{unbiased}$ = -0.13).

To investigate whether the differences between the two groups are meaningful, a visualization was made from the analysis to accompany the mean difference estimation on the mean duration of pause, which has the largest effect size. This information is presented in Figure 7.13.

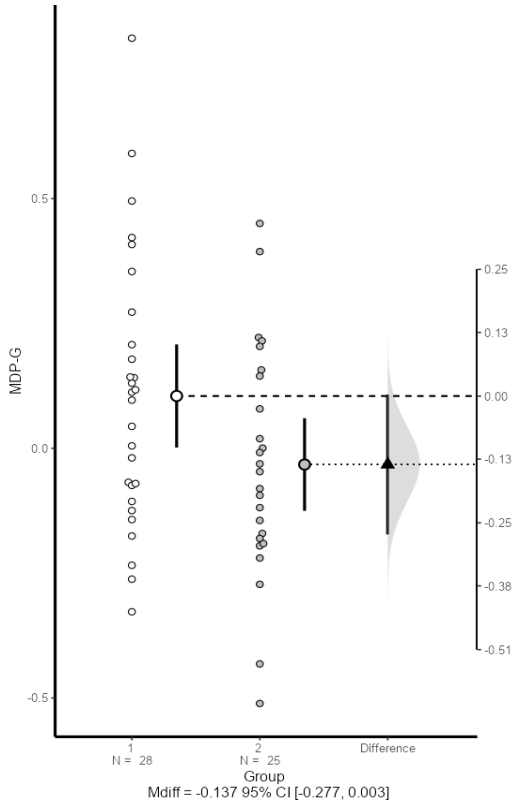


Figure 7.13. Scatter plot of means and 95% confidence intervals (CIs) for the gain in mean duration of pause (MDP-G) score for the two groups

Note: 1. The difference between the group means, with its 95% CI, is shown on a floating difference axis at the right. 2. The explicit group is 1, and the implicit group is 2.

It can be seen from Figure 7.13 that the CIs of the two groups overlap for more than half of one arm of the CI, which represents the average margin of error (MOE). Therefore, the evidence of a difference between the two means is not meaningful. Other comparisons with smaller effect sizes also produce more overlapping CIs (see other figures in Appendix 7).

Since the differences between the two groups in speech fluency gains are not statistically significant and not meaningful, the effects of both instructional conditions on EFL learners' speech fluency development are considered comparable. As suggested by the effect sizes, each instructional condition seems to have a better impact on different speech fluency aspects.

The impact of explicit instruction is more visible than the implicit instruction in three speech fluency measures: the phonation time ratio, mean length of run, and mean duration of pauses. The effect sizes of the difference in these three measures are around the medium. It seems that when participants in this group were asked to pay attention to their frequent pausing, use lexical fillers during pausing or instead of pausing, and avoid repetition and repair moves while reporting a speech they could speak longer between the pauses (increasing the mean length of run) (see Chapter 6 for more details on the training strategies).The pauses they produced between utterances (run) also improved in terms of the duration (improved mean duration of pauses).

As reported in Table 7.2 above, the participants in the EG produced the same number of pauses in the pre-test and post-test, but the duration of these pauses was shorter in the post-test. Therefore, the explicit strategy training instruction could impact learners' speech fluency since they could produce a longer utterance between shorter pauses. The results of these three measures also indicate that the strategy training instruction might influence learners to proceduralize their language knowledge (lexical and syntactic) in speech production (see Section 7.3.1.1).

Another implication that could be derived from the results is that, for the articulation rate, the implicit instruction had a better impact than the explicit instruction. An improvement in the articulation rate indicates an improved speed of speech. Therefore, providing learners with practice opportunities only (repeating practice) could enable them to improve the speed of their speech.

Learners in this IG were not given the fluency strategy training, so they were considered unaware of any strategies that could be used to improve their speech fluency. Nevertheless, these learners could still produce faster speech. However, when learners spoke more quickly, this seemed to affect the frequency and duration of pauses in their speech. The descriptive analysis performed for this IG (see Table 7.3) revealed that the participants produced more pauses with longer duration in the post-test. More and longer pauses created by learners when producing a faster speech could indicate that they were experimenting with new automatized language during speech production.

Hence, the explicit instruction with repeated practice opportunities also impacted learners' speech fluency in faster speech and, potentially, their automatized language production. However, as de Jong and Perfetti (2011) argue, improvement in terms of the articulation rate is not an indication of the proceduralization of lexical and syntactic knowledge.

Regarding the comparison of oral proficiency gains between groups, the same analyses were also applied. Bar charts were used as the first analysis to enable a visual inspection of the gains (Figure 7.14).

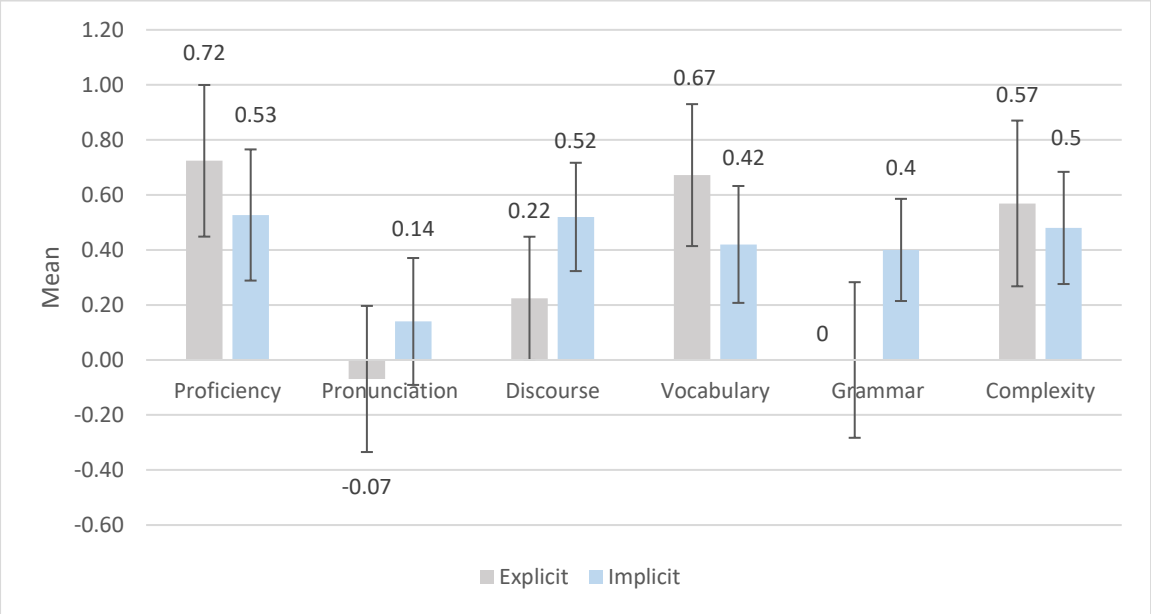


Figure 7.14. Both groups' oral proficiency gains

Note: 1 denotes the explicit group, and 2 is the implicit group.
Error bars represent 95% confidence intervals of the means.

From the six scales detailed in Table 7.14, the IG has a higher gain than the EG for three scales: pronunciation, discourse, and grammar. The EG experienced a setback in their pronunciation, while there was no gain for their grammar since the mean was zero. The highest difference in gain between the two groups was also found for grammar. On the other hand, the EG's gains were higher than the IG in terms of the other three scales: general proficiency, vocabulary, and complexity. The gain in vocabulary was the greatest difference between the groups.

Further analysis using the Mann-Whitney U Test was then performed to investigate the level of significance of the differences shown by the bar chart in Figure 7.14. The results of the Mann-Whitney U Test are presented in Table 7.16.

Table 7.16. The results of Mann-Whitney U Test of oral proficiency gain for both groups

Scales	Groups	N	Mean Rank	*z*	*Sig. (p)*	*Observed Power*
Proficiency	Explicit	28	30.21	-1.41	0.16	
	Implicit	25	24.36			

	Total	53				
Pronunciation	Explicit	28	26.12	-0.72	0.47	
	Implicit	25	29.10			
	Total	53				
Discourse	Explicit	28	24.59	-1.55	0.12	
	Implicit	25	30.88			
	Total	53				
Vocabulary	Explicit	28	31.48	-2.08	0.04	0.32
	Implicit	25	22.88			
	Total	53				
Grammar	Explicit	28	23.69	-2.03	0.04	0.68
	Implicit	25	31.92			
	Total	53				
Complexity	Explicit	28	30.10	-1.39	0.16	
	Implicit	25	24.48			
	Total	53				

Note: Observed power is provided only when the result is significant.

As reported in Table 7.16, the gain differences between the two instructional groups reached a statistical significance in two scales: vocabulary (z = -2.08, p = 0.04) and grammar (z = -2.03, p = 0.04). However, the observed power for these two scales (0.32 for vocabulary and 0.68 for grammar) is lower than 0.8, which is the acceptable value. Therefore, while these results are statistically, they are considered inconclusive. This may offer further support for the view that the effect of the explicit strategy training instruction is comparable to the effect of implicit instruction in improving EFL learners' oral proficiency.

In order to determine the magnitude of the difference between both groups' oral proficiency gains, estimation analysis was performed on the six oral proficiency scales (see Table 7.17).

Table 7.17 The results of independent mean difference analysis on raters' judgement for gains, a comparison between the explicit and implicit groups

Measure	Group	M	95 % CI		SD	$d_{unbiased}$	N
			Lower	Upper			
Proficiency	Implicit	0.56	0.35	0.77	0.53		25
	Explicit	0.72	0.47	0.98	0.69		29
	Difference	-0.16	-0.50	0.18	0.62	-0.26	54
Pronunciation	Implicit	0.14	-0.11	0.39	0.62		25

	Explicit	-0.07	-0.32	0.18	0.66		29
	Difference	0.21	-0.14	0.56	0.65	0.32	54
Discourse	Implicit	0.52	0.31	0.73	0.53		25
	Explicit	0.22	0.02	0.43	0.56		29
	Difference	0.30	0.00	0.60	0.55	0.53	54
Vocabulary	Implicit	0.42	0.19	0.65	0.57		25
	Explicit	0.67	0.43	0.91	0.65		29
	Difference	-0.25	-0.59	0.08	0.61	-0.41	54
Grammar	Implicit	0.40	0.20	0.60	0.50		25
	Explicit	0.00	-0.26	0.26	0.71		29
	Difference	0.40	0.06	0.74	0.62	0.64	54
Complexity	Implicit	0.48	0.26	0.70	0.55		25
	Explicit	0.57	0.29	0.85	0.75		29
	Difference	-0.09	-0.45	0.28	0.67	-0.13	54

Note: 1. The standardized effect size is d_{unbiased} because the denominator used was SDpooled. The standardized effect size has been corrected for bias. The bias-corrected version of Cohen's d is sometimes also called Hedges' g. 2. Negative mean scores in a group indicate that the scores in the post-test are lower than the pre-test.

Table 7.17 shows that the IG has higher means than the EG for three scales: pronunciation (M_{diff} = -0.21, [-0.14, 0.56]), discourse (M_{diff} = 0.30, [0.00, 0.60]), and grammar (M_{diff} = 0.40 [0.06, 0.74]). The effect sizes of these three scales range from small to medium ($d_{unbiased}$ = 0.32 for pronunciation, $d_{unbiased}$ = 0.53 for discourse, and $d_{unbiased}$ = 0.64 for grammar). Meanwhile, the EG has higher means than the IG for the other three scales: proficiency, vocabulary, and complexity, although the effect sizes for these are small ($d_{unbiased}$ = -0.26 for proficiency, $d_{unbiased}$ = -0.41 for vocabulary, and $d_{unbiased}$ = -0.13 for complexity). Further investigation to determine the meaningfulness of the difference using a scatter plot was performed for grammar (Figure 7.15), which has the largest effect size in the analysis.

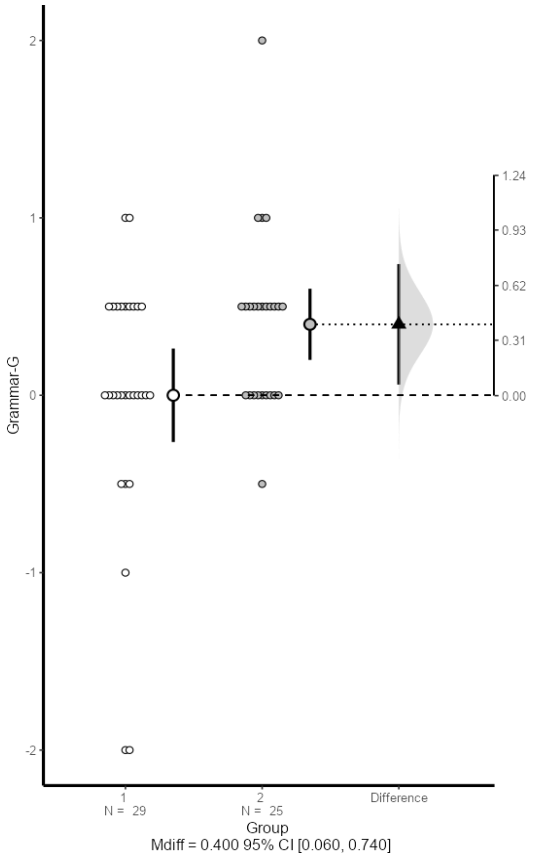


Figure 7.15. Scatter plot of means and 95% confidence intervals (CIs) for the gain in grammar scale for the two groups

Note: 1. The difference between the group means, with its 95% CI, is shown on a floating difference axis at the right. 2. The explicit group is 1, and the implicit group is 2.

As shown in Figure 7.15, the CIs of the two groups overlap for more than half of one arm, indicating the difference between the two means is not meaningful. These results show that when measured based on the raters' judgement, both instructional conditions also have a comparable impact on learners' oral proficiency.

The explicit strategy training instruction seems to have a better impact on learners' general proficiency, vocabulary, and complexity. It seems that instructing participants to explicitly focus on aspects that could improve their speech fluency (such as avoiding and filling their pauses, and avoiding making repetitions and repair moves), could improve their oral performance by producing a better organized speech with varied vocabulary and some embedded clauses.

Meanwhile, the impact of implicit instruction on learners' pronunciation, discourse, and grammar seems to be better than explicit instruction. Therefore, it can also be argued that implicit instruction could improve learners' oral performance slightly differently. When instructed implicitly, learners were able to produce a more plausible and logically informed speech. The speech also had fewer phonemic and grammatical errors and better intonation patterns.

7.4 Summary

In this chapter the findings have been presented for the strength of the relationships (considered an effect) of explicit strategy training and implicit instruction on the participants' speech fluency and perceived oral proficiency. For the participants within the study context, these two instructional conditions seemed to have a comparable impact on developing their speech fluency and oral proficiency.

Each form of instruction had some kind of impact on rather different aspects of speech fluency. The explicit strategy training's impact on participants' phonation time ratio and mean duration of pauses was stronger than the implicit strategy training's impact. The impact of implicit instruction was stronger than explicit instruction on the articulation rate. Meanwhile, the impact of both instructional conditions on the speech rate was highly comparable, being quite similar.

Although not statistically significant, for the explicit instruction, medium effect sizes were noted in improving the phonation time ratio and the mean duration of pauses. The effect sizes of other variables were small. The explicit instruction also seemed to impact learners' proceduralization of language knowledge. On the other hand, the non-significant results in the implicit instruction were accompanied by small effect sizes on the improvement of the articulation rate and the mean length of run, with the effect sizes of other variables being negligible. The implicit instruction had no impact on developing learners' proceduralization.

In addition, this chapter also included the results of the raters' judgement on participants' oral proficiency. Using holistic and analytic rubrics, the two instructional conditions seemed

to result in significant and meaningful improvements according to the raters' scores. Each instructional condition also had a stronger impact on different aspects of oral proficiency.

The impact of the explicit strategy training instruction was stronger than that of the implicit instruction on learners' general proficiency, vocabulary, and complexity. On the other hand, the impact of implicit instruction was stronger on learners' pronunciation, discourse, and grammar. Individually, each form of instruction was found to have a meaningful impact on oral proficiency aspects. Meanwhile, learners' proficiency, motivation, and self-regulation did not correlate with speech fluency development. The theoretical and pedagogical implications of these results for EFL and ESL learning are discussed in Chapter 9.

Chapter 8: CDS Phase Transition Analysis

In this chapter, the results of phase transition analysis conducted on the selected participants' data are discussed. The chapter begins with Section 8.1, in which the procedures of selecting the participants are described. Section 8.2 provides details of the phase transition analysis, including the procedures applied. The phase transition analysis consisted of three criteria: sudden jump, anomalous variance, and qualitative change (see Chapter 5 Section 5.2.2 and Chapter 6 Section 6.6.2.2). Finally, a summary regarding the findings of the phase transition analysis is given in Section 8.3.

8.1 Data Selection Procedures

As described in Chapter 6, the data to be included in the phase transition analysis were taken from four participants: two participants with high speech fluency gains from each group were selected. The speech fluency gains were assessed by comparing the pre-test and post-test scores of four speech fluency measures: the speech rate, phonation time ratio, mean length of runs, and mean duration of pauses. Gains in each of the first three were represented by higher post-test scores. For the mean duration of pauses, these gains were represented by lower post-test scores since this measure dealt with the length and number of pauses. From a complex dynamic system perspective, high gains in speech fluency could be considered changes in speech fluency, which was regarded as the system being investigated.

For selection, the mean scores of the four measures in the pre-test and post-test were generated for all participants in both groups by using the case summary tool of the SPSS program. Then, the results were moved into a Microsoft Excel spreadsheet for further calculation. In the spreadsheet, the mean scores of the pre-tests of each measure were subtracted from their respective post-test scores to ascertain the difference or gain. Next, these mean differences were ranked by using the formula available in the spreadsheet. The resulted rank scores were then equated to achieve an average rank score. First, the scores from speech rate, phonation time ratio, and mean length of run were added together. The

result was then subtracted by the mean duration of pauses' score. As detailed in Table 8.1, the results of the equation took the form of average gain scores for each participant in both groups. The two participants with the highest scores from both groups were then selected for the analysis. Since there were two participants in second place with similar ranked scores (63) in the implicit instruction group (IG), the student in the higher position was chosen. Thus, in the IG, Student 8 and Student 22 were selected. In the explicit instruction group (EG), Student 20 and Student 18 gained the highest scores and were chosen. Interestingly, the two participants representing each group were in different proficiency levels. Student 20 and Student 8 were classified as high proficiency participants with TOEIC scores of 635 and 655, respectively. On the other hand, Student 18 and Student 22, whose TOEIC scores were 300 and 390, respectively, were low proficiency participants.

Table 8.1 The results of the ranking of all participants

	Explicit Group		**Implicit Group**	
No	**Participant**	**Score**	**Participant**	**Score**
1	Student 20	78	Student 8	65
2	Student 18	77	Student 22	63
3	Student 14	74	Student 24	63
4	Student 3	61	Student 11	60
5	Student 25	56	Student 10	57
6	Student1	55	Student13	53
7	Student 12	53	Student 19	41
8	Student 2	52	Student 6	39
9	Student 24	40.5	Student 14	33
10	Student 27	35	Student 25	30
11	Student 7	32	Student 4	29
12	Student 11	31.5	Student 17	29
13	Student 5	29	Student 9	25
14	Student 21	28	Student 3	24
15	Student 4	26.5	Student 1	21
16	Student 23	26	Student 18	19

Explicit Group			Implicit Group	
No	**Participant**	**Score**	**Participant**	**Score**
17	Student 6	22.5	Student 2	18
18	Student 16	22	Student 7	10
19	Student 19	22	Student 20	8
20	Student 9	21	Student 5	3
21	Student 15	16.5	Student 12	1
22	Student 8	8	Student 16	-1.5
23	Student 28	8	Student 23	-6
24	Student 22	6.5	Student 15	-14
25	Student 17	4.5	Student 21	-19.5
26	Student 10	2		
27	Student 29	-5		
28	Student 13	-6		
29	Student26	-6.5		

Note. Scores represent means for differences in the speech rate, phonation time ratio, mean length of run, and mean duration of pauses when comparing the pre-test and post-test.

8.2 Phase Transition Analysis

The phase transition analysis involved time series data of the four participants. This data included the pre-test, post-test, and eight task performances. Therefore, the data consisted of ten series. The analysis involved three criteria: (1) a sudden jump, (2) anomalous variance, and (3) and a qualitative change in the attractor. The sudden jump was investigated by using change point analysis and moving windows of the min-max graph. "The change point analysis indicates the points at which a sudden change occurs"(Baba & Nitta, 2014, p. 14). The change point analysis was performed with the Change Point Analyser (Taylor, 2000). The change point analyzer displays the analysis results in figures and tables containing information about the change. Therefore, the results could be visually inspected.

This change point analysis was followed by a moving min-max graph (Van Geert & Van Dijk, 2002; Verspoor et al., 2011) to examine the identified change points further. The moving

min-max graph was also used to accentuate the size of fluctuations within moving time frames (Verspoor et al., 2011). A min-max graph presents the data being analyzed in a bandwidth of scores instead of lines. The graph was created by a moving window concept in which the time series data were divided into smaller timeframes according to the available data points. Since this study only consisted of ten data points, each window only contained three data points. In the application, to use the words of experts, "each window's timeframe moved up one position, and each window partly overlaps the preceding one, using all the same measurement occasions minus the first and plus the next" (Van Geert & Van Dijk, 2002, p. 354).

The first window in this data consisted of data points from Week 1 to Week 3, the second window contained the data from Week 2 to Week 4, and so on. Then, each window's maximum and minimum values were calculated and converted into a line. In so doing, the moving min-max graph would be able to display the variability of general developmental trends in the time series data (Van Geert & Van Dijk, 2002).

The anomalous variance was analyzed by establishing the variability in the time series data. For this purpose, a line graph of the moving standard deviation of residuals (window size = 3) and a line graph of the absolute values of the analyzed variable were used. The residuals' standard deviations were preferred to avoid the augmentation of the moving standard deviation due to the increase in raw value (Baba & Nitta, 2014; Van Geert & Van Dijk, 2002). The line graph containing the moving standard deviation residuals and the absolute values data was used to visualize the extent of fluctuations in the development of the analyzed variable.

An increase in variability is represented by higher fluctuations in the magnitude between the two lines in the graph. Increasing variability is indicative of the system's instability. "A system is supposed to be relatively stable over some initial period of time. This period must be followed by a period in which this system becomes unstable, which results in large variability" (Van Geert & Van Dijk, 2002, p. 363). If this variability happened near the vicinity of a phase transition, it would confirm the occurrence of a sudden jump.

Meanwhile, to find indications of a qualitative change in the four participants' speech performances, the transcription of these was used. The four participants' speech samples, both in the original and pruned transcription, were analyzed. Investigation of the change in the original transcriptions involved two variables: pauses and repetition. These were chosen because these two variables showed a visually distinctive change from the beginning to the end of the data series. Information regarding pauses and repetition was also expected to reveal participants' speech characteristics before and after the change. In the investigation, these two variables were divided by the number of syllables of the pruned transcriptions to obtain ratios for analysis purposes.

Investigation regarding the change in pruned transcriptions was based on the change in the speech's contents and composition. In terms of contents, participants' transcripts were compared with the news video transcripts to find the similarity of the informational contents. Hence, this content analysis would provide information regarding how the participants absorbed and applied the language input into their speech.

The investigation on compositions focused on how the participants conveyed the news following the standard composition of a news presentation. The composition consists of an introduction, body, and conclusion. News items on television, radio, or in newspapers usually start with a presentation of opening information or an introduction of the news to audiences. Usually, this introduction highlights the details in the news being presented in one or two utterances, following the classic 5 W's (what, who, where, when, and why) questions and an H (how). For example, the introduction in the news about accidents usually informs audiences about what happened, who was involved, and where and when it happened. The introduction also advises why or how the accident happened. Then, the body of the news discusses each aspect of the information in the 5 W and the H questions in detail.

For news videos generally seen on television, the body of the news contains interviews with the people who are directly or indirectly involved in accidents, such as the victims, rescuers, police or security officers, or experts. In the end, the news is usually finalized by concluding

information, which often does not function as a conclusion but as supplementary information that increases the 'value' of the news.

Although this composition is not a strict rule that all news items should follow, it can be found in daily television or radio news reports. For example, one news video used as the material in the study contained information about an airplane accident in Nepal. In the introduction, the presenter announced that a small commercial airplane from Kathmandu to Lukla (a small town near Mount Everest) plunged only minutes after take-off. The introduction continued with mentioning the number of victims. In the body, information from a police spokesperson interviewed at the accident location was conveyed. The body of the news also contained information regarding some similar accidents that happened previously in the region. Then, the news item was completed with a concluding remark informing the audiences that tourists continued using the small airplanes despite the frequent accidents.

Most of the news videos used in the study followed a similar composition structure. Therefore, knowing this would help trace the similarity between participants' speech samples and the news. The investigation on composition was also targeted at finding specific utterances which the participants borrowed from the news videos and used in their speech. Hence, the analysis could reveal developmental changes in participants' speech. It was also expected that the analysis would provide information regarding the kind of phrases that could easily be absorbed and applied by L2 learners.

Ten data series taken from Week 1 to Week 10 of the study were used for analysis purposes. The variables investigated were the speech rate, phonation time ratio, mean length of run, mean duration of pause, and number of syllables. Although the number of syllables produced by a participant in a specific time duration does not directly represent a distinct and discrete measure of fluency, the role of this measure cannot be neglected because it is part of the system being investigated (speech fluency). "From the CDS perspective, the object to be explored is a system and not a variable in the traditional sense. A system is inevitably multidimensional, which cannot be simplified as one measure" (Baba

& Nitta, 2014, p. 13). Besides, to the author's knowledge, this study is the very first one investigating speech fluency by using the CDS perspective; therefore, it was exploratory in nature in terms of the included variables, or, to use a CDS term, the focus of observation. As a comparison, Baba and Nitta (2014) used the number of words (produced by participants in ten minutes) as their focus to observe L2 writing fluency.

8.2.1 The results from Student 20 (Explicit Group)

8.2.1.1 Sudden Jump

The results of the change point analysis of Student 20's data revealed that only the number of syllables showed an indication of a sudden jump (phase transition). For this variable, one change point (jump) occurred during the time series data. As shown in Figure 8.1 and Table 8.2, this change occurred in Week 7, with a 95% confidence level. The confidence interval fell between Week 7 and 10. The change point location was Week 7. Therefore, Week 1 to Week 6 was considered the first phase in the development, while Week 7 to Week 10 became the second phase. Before the change, Student 20 produced an average of 108 syllables. The average number of syllables increased to 213.25 after the change. The nomenclature of Level 1 was ascribed to this change, indicating that it was the first to be detected in the analysis as it was the most visually apparent in the plot.

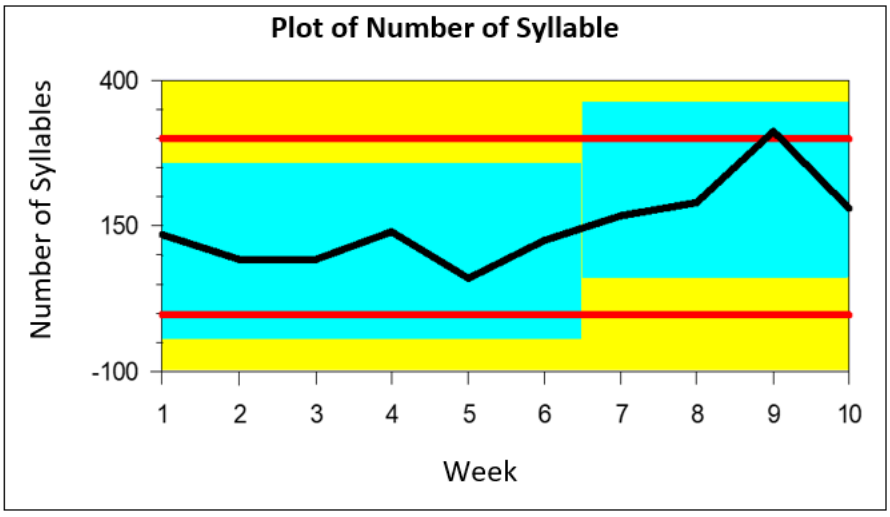


Figure 8.1 Change point analyzer plot of number of syllables produced by Student 20

Notes: 1. The red lines are control limits representing the maximum range that the values are expected to vary, assuming no change has occurred.
2. Shifts in the blue region represent changes in the time series data

Table 8.2 Change point analysis result for number of syllables for Student 20

Detected Change	Confidence Interval	Confidence Level	Average Change of Number of Syllables		Level of Change
Week	Week		From	To	
7	(7, 10)	95%	108	213.25	1

Notes. 1. The number of syllables is the actual number of syllables produced in the pruned transcription.
2. The average change is based on the average number of syllables in each phase (the first phase consists of Week 1 to Week 6, and the second phase is from Week 7 to Week 10).

Figure 8.2, which is a moving min-max graph, confirms that the trajectory of the participant's number of syllables showed periods of progress and regress, rather than a linear developmental path. The development seemed to be carried out in two noticeable phases. The first phase happened during Week 1 to Week 6, and the second was from Week 7 to Week 10. During the first phase, the number of syllables increased steadily. The highest value occurred in Week 4 and then this dropped in Week 5. Although it slightly increased in Week 6, the value was still lower than that of Week 4. The bandwidth between the number of syllables and its maximum value seemed to be stagnant. In Week 4, the bandwidth between the variable and its minimum value enlarged before closing down in Week 5.

The second phase was marked by a sharp increase in the number of syllables, with Week 9 having the highest. The bandwidth also grew larger during this period, especially between Week 6 and Week 8. The bandwidth ended in Week 9 because the moving window's calculation stopped at this point. The results shown in this moving min-max graph confirmed the change point analysis result, which was that a significant change in the time series data occurred during the second phase (after Week 6).

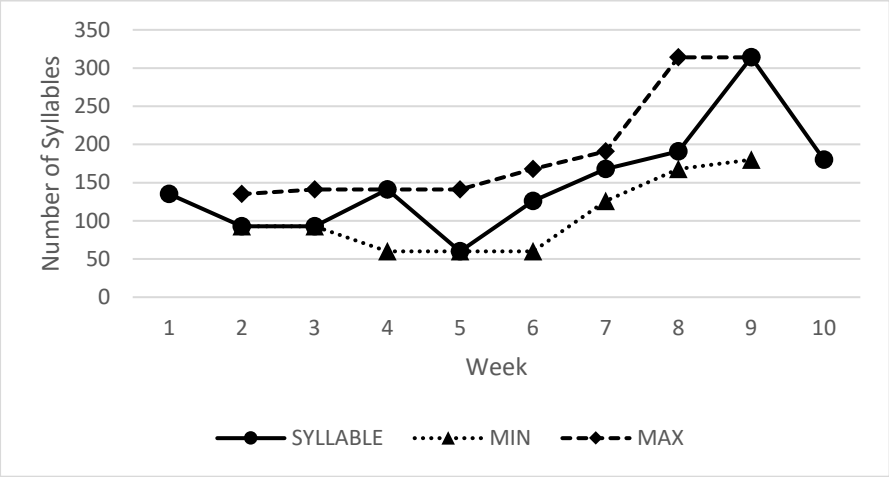


Figure 8.2 Change in the number of syllables of Student 20 as indicated by a moving min-max graph

A Monte Carlo analysis with 10,000 replications (van Geert, Steenbeek, & Kunnen, 2012) was used to test whether there was a statistical difference between the two phases of development. A conventional significance level (p = .05) was applied in the analysis. The analysis showed that the number of syllables produced in the second phase was significantly higher than in the first phase (p = 0.010). The result also corroborated the division of the learning period into two development phases revealed by the min-max graph.

8.2.1.2 Anomalous Variance

To investigate the anomalous variance, moving standard deviation residuals of each moving window were analyzed together with the absolute values of the number of syllables. The number of syllables variable was taken to be the total number of syllables counted based on the pruned transcription of the participant's speech performance on Task 3 (see Section 6.5.1.3 for details). Figure 8.3 shows the analysis results in the form of a line graph. The up/down bars represent the magnitude between the moving standard deviation residuals and their absolute values. As can be seen, the bars show some fluctuations from Week 2 to Week 6. A high fluctuation seems to occur between Week 4 and Week 6, in which the smallest magnitude is seen in Week 5. On the contrary, the magnitude of differences seems to increase steadily from Week 7 to Week 9, with very low fluctuations, as indicated by the

up/down bars. Therefore, the high fluctuations that occurred during Week 4 and Week 6 could be considered critical fluctuation or large variability, indicating that the system was preparing to enter a phase transition. This result supports the result of the change point analysis above, which suggests a sudden jump occurred in Week 7.

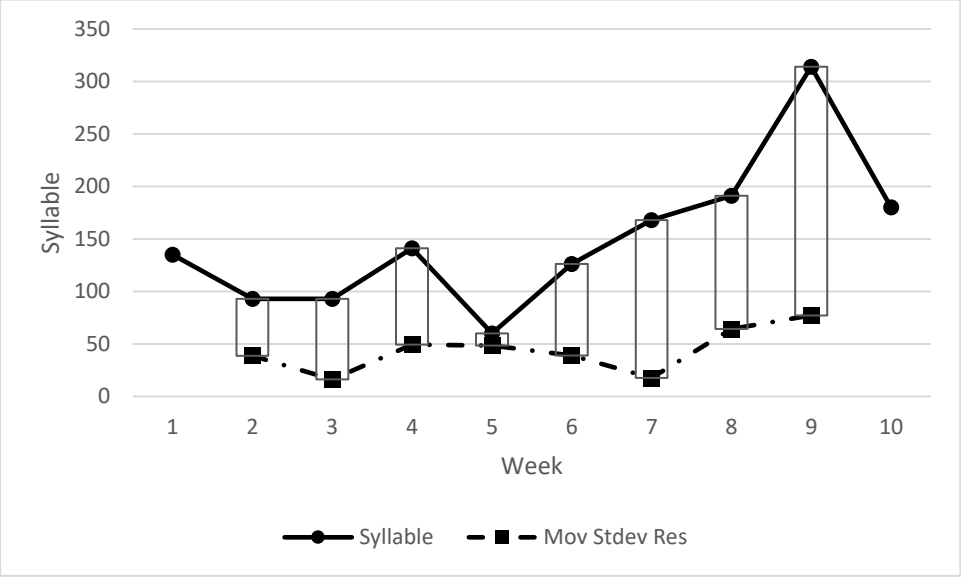


Figure 8.3 Magnitude of change in the number of syllables of Student 20

Note. The up/down bars represent the magnitude between the moving standard deviation residuals and their absolute values.

8.2.1.3 Qualitative changes in the attractor

Student 20's speech production seemed to change qualitatively after the first phase transition or when the development entered an attractor state. As discussed in Chapter 5 Section 5.2.1.6, the attractor state is the state that occurs when systems remain in the same condition for a period of time as a result of the development process (De Bot & Larsen-Freeman, 2011). Systems remain in their current attractor state until another change driven by substantial energy occurs and causes them to move into another state of development or a new attractor state. Therefore, qualitative changes in the attractor can serve as evidence of the systems' reorganization, which qualitatively differentiates a current performance mode from a previous one.

Changes in pauses and repetition ratios

The first change to be investigated qualitatively was the change in how the participant paused and made verbatim repetitions in her speech. In this investigation, the change point analyzer tool was not used since it was part of the qualitative analysis to support the change point analysis performed previously. Therefore, the change mentioned in this analysis was not meant as an indication of a sudden jump in the variables' development but merely as information on how the variables changed over time. Here, Student 20's speech samples indicated a change in pauses and repetition ratios. Figure 8.4 shows the change in ten weeks or between the two phases revealed by the change point analysis.

For both variables, different patterns following the two development phases are displayed. Pause ratios were fluctuating during the first phase, ranging from 15% to 21%. The trajectory of the ratio increased at the end of the first phase (Week 6) and then gradually decreased towards the end of the second phase. This gradual decrease in the number of pauses might indicate that the system entered its attractor state. The polynomial trend line in the pause ratios decreased, indicating that the participant's pauses decreased over time. This result indicates the participant's increased speech fluency since she was able to produce more syllables in her speeches while at the same time reducing the number of pauses.

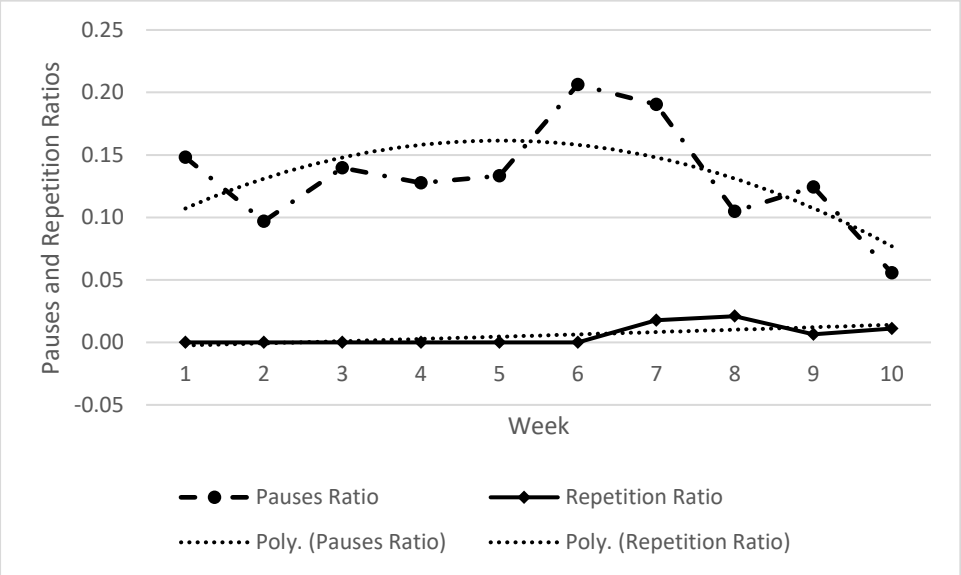


Figure 8.4 Ratios of pauses and repetitions against the number of syllables of Student 20 with the polynomial trend line (2nd degree)

Note. The ratios of pauses and repetitions are the result of dividing the variables by the number of syllables in the pruned transcriptions.

The repetition ratios were stable during the first phase, in which the participant did not make any repetition in the speech. Then at the beginning of the second phase in Week 7, the ratio suddenly increased. The participant continued to make repetitions in her speeches at the data points which followed towards the end of the phase, indicating that the development entered an attractor state after a phase transition. The polynomial trend line for repetition shows an increase, indicating that the participant's repetitions increased over time. Therefore, Student 20 tended to create more repetitions in longer speeches when more syllables were produced.

Changes in contents and compositions of the speech

Contents. It was found that Student 20 did not cover all segments of information in each report during the first phase. For example, the participant only covered two thirds of the news information in her report in Week 1. The news video used in that week provided five information segments regarding a terror attack in Jakarta in January 2016. The first segment contained information regarding a Friday prayer sermon that condemned the attack. The second segment presented responses from three interviewees who also condemned the attack. The interviewees stated that Islam had nothing to do with the attack. They also denied a claim which connected the terror attack with Islamic teaching. The third segment comprised information about an inter-faith dialogue of peace held by Nahdatul Ulama, a prominent Muslim scholars' organization in Indonesia. In the fourth segment, a video campaign released by Nahdatul Ulama a year before was described. This was a campaign to fight extremist ideology. The last segment was a summary of the terror attack. As can be seen in the following transcription, the student only reported three segments: the second, third, and the last out of these five. Interestingly, the participant started the report with the last segment from the video, indicating that she knew that she could use the segment as an introduction for the report:

> *Ok guys, here I am reporting from Indonesia tragedy. It is about bomb attack in Indonesia. It is the first bomb attack by terrorist that the suspected by Moslem. (5)*
>
> *But there are some of Moslem said and rejected that there is no way there is no reason Moslem said be a terrorist and terrorist also does not mean Moslem. Also there is some of Moslem said that they have to believe with the government to reduce the bomb the terrorist and about the bomb attack. (2)*
>
> *And Nahdatul Ulama has said that they have forty million Moslem recording to the data. (3)*
>
> *Thank you. (Student 20)*

Similar performances were found in reports for Weeks 2 and 3, in which the participant only conveyed around two or three segments out of the five available. In Week 4, the participant covered most of the information from the news. However, the speech was relatively short in terms of duration and the number of syllables it contained. In Week 5, the student's speech was even shorter. The student only reported a summary of two segments of information from the five segments in the news video. There were no details added to the summary; therefore, this time, the duration and number of syllables of the speech turned out to be the lowest.

In contrast, during the second phase, Student 20's reports covered most of the information, including details. Although there was no detailed information in Week 6's report, the participant began to cover more information in Weeks 7, 8, 9, and 10. In Week 9, in which the participant produced the highest number of syllables, the student reported a news video about hate crimes in Canada. The student mentioned it as "our city," replicating the news reporter in the video. The news video contained four segments of information. As is common in news, the first segment highlighted the entire information in the news. The second one presented information regarding a hate crime incident that happened in Ontario involving a white man and two India-born Canadians who were all fighting over a parking space. The third one was about a previous hate crime in Toronto between a white man and a Muslim family in front of a shopping centre. The fourth was a report about a video message from a legal expert who gave his opinion regarding the action and consequences of hate crimes. The legal expert also stated that the two incidents were hate crimes. In her

speech, Student 20 reported all of the information sequentially, and in detail, as can be seen from the following transcription:

> *I am Shiva; I am reporting about hate crime right now. Nowadays, hate crimes widespread into our society and it becomes our concern. And today I will tell you there are two cases about hate crimes. (1)*
>
> *The first case happened, I am sorry, happened in parking space and then the man and his wife get become a victim by the driver. The driver was a citizen in the city. And also there are a fight about the parking space. The driver, the real citizen talk and speech about the hate about the treat them. And then, the man's wife record the video. And this cases, the first case is now the concern by the judge. And the judge will decide it is a hate crime or not. (2)*
>
> *And the second case happened, I am sorry, happened at the front of some building at a city. And the real citizen, the resident or the people who grow up in that city talk hate speech also with the Muslim citizen right there. And the judge talked that is a hate crime, hate criminals. (3)*
>
> *And also, there is a crime defence lawyer; Mark Sanders said that the hate is a motivation to do the crime and we will have to concern about it and aware about the hate crime because it is really dangerous for us and for our society. And for punishment of the hate crime is a lifetime. (4) (Student 20)*

Compositions. The participant's compositions changed considerably from the first phase to the second phase, which could be found in the body of the composition while the introductions and conclusions remained the same. In addition, the participant also started borrowing some phrases from the news. In general, the participant produced a similar introduction in all performances across the two phases. Usually, this began with her introducing herself by mentioning her first name (full name in one performance) and then a brief outline about what happened in the news.

Here, the participant did not follow the general guidelines of a news introduction, as discussed previously (see Section 8.2). In most performances, the participant mentioned the place of the accident or incident (the 'where') but never included all the 5W's and H. For example, in Week 6 she reported the date and the kind of accident in the introduction. However, she did not mention its cause. Some examples of introductions from Week 5, Week 6, and Week 10 reports are presented below.

Introduction in Week 5:

"Here I am, Shiva, now reporting about the Nepal airplane crash."

Introduction in Week 6:

"Wildfire hit Sumatra Island Indonesia on June eighteen, two thousand thirteen."

Introduction in Week 10:

"I am Shiva Nasila reporting about earthquake in Indonesia." (Student 20)

In terms of the conclusion, the participant did not include any supplementary information, as provided by most of the news videos. She only gave a standard, closing remark, "*thank you*". The participant appeared to be more consistent in making this remark in the second phase, where she created four remarks in Week 7 to Week 10. During the first phase, this remark was only found in Week 1. In Week 4, the participant produced a different concluding remark by saying, "*I am Shiva reporting from the seashore, Russian."* However, this was not supplementary information (see Section 8.2).

Meanwhile, the change in the body of the compositions in the second phase was marked by complete information regarding answering the 5 W's and H questions in detail. The participant also presented the report more sequentially. In contrast, the body of the reports in the first phase showed a lack of detailed information. A comparison of participant's performances between Week 4 and Week 7 shows this. Both of the reports in these weeks dealt with transportation accidents involving an airplane (Week 4) and a ship (Week 7). As can be seen in the following transcription, the participant reported the accident in terms of what happened, who were involved, and where it happened in Week 4. However, she did not mention why and how the accident happened.

> *This airplane kill all the climbers. And well, the climbers are going to the Everest Mountain. There are sixteen passengers that have been killed. And also the airplane have been crashed after fifteen minutes take-off. (Student 20)*

In Week 7 (see the following transcription), the participant reported all the detailed information regarding what happened, when and where it happened, who the victims were, and how the accident occurred:

> *The Thomson cruises will go to Spain to Portugal, and before that trip, the cruise ship and the crews member will held the rescue personal training. This has been happened before they go to the Spain, I am sorry, they go to the Portugal, and when they have the rescue personal training the crews personal member go to the lifeboat, but unfortunately, the cable of lifeboat was broke, and the lifeboat fall down and also the people who top on them. So on the cruise ship, there are fifteen hundred passengers. And three crews member injured. This happen the lifeboat fall down until twenty metres into the ocean. (Student 20)*

Similar conditions were found in another comparison between Week 2 (a terror incident) and Week 8 (domestic violence). In Week 2, the participant reported a terror incident in France. However, detailed information concerning the actual process of the incident was not given by the participant. As can be seen in the following transcription, the participant only talked about the victim without specifying what actually happened, when it happened, or why and how:

> *As the suspected with a shot for policemen and one person killed there. The suspected also ISIS terrorists in the supermarket. And now we know that a few years ago France also have two attack in the two thousand fifteen and also two attack bombers on two thousand sixteen. (Student 20)*

In contrast, the participant reported more detailed information in Week 8. This time, she reported a domestic violence incident involving a woman who later became a domestic violence activist. The news video was about a benefit luncheon where the woman, a former news anchor in the city, was the main speaker. The woman talked about her experience as a former victim of domestic violence. In the report, as can be seen in the transcription below, the participant reported the story from the beginning to the end, following the sequence in the news video. She detailed what happened to the woman regarding the domestic violence she experienced. First, the participant mentioned the woman as a violence survivor. Then she explained how and when the violence started. After that, she informed about how the woman got out of the situation. Finally, similarly to the news video, the participant also included information regarding the woman's activism at the end of the report.

> *She is a violent survivor in the case of domestic and family violence in that city. It began when she got married in two thousand and ten, and the rest of her marriage with her*

> *husband only a year until she had a son. When her son in five-week, her husband changing directly into a rough person. He did bad thing on her, and fortunately, her neighbour heard her screaming at the time and directly called the police. And now she help another victims of violence and gave her encourage speech with the others. She said that save your family, save your children, and become be brave and speak up with your personality with your experiencing about the violence. (Student 20)*

In terms of borrowed utterances, the participant's speech also showed a substantial change. In the first phase, she only used one utterance from the news video in one of her speeches. This was "*truck lost control,*" reported in Week 3, which was the same phrase used in the news video. As time progressed, borrowed utterances could be found in four weeks (Weeks 7, 8, 9, and 10) during the second phase. In Week 7, the participants used the utterance "*fifteen hundred passengers*" as initially stated in the news video. This is particularly significant as usually, L2 students in Indonesia would prefer to say "*one thousand five hundred*" instead of "*fifteen hundred*." In Week 8, the participant used the utterance "*save your family! Save your children!"* The actual utterance in the news videos was "*save yourself! Save your family!"* Although this was not precisely the same as the original, it was still considered borrowing since the actual meaning and how the sentences were reported in the speech were similar. In Week 9, the participant used one specific utterance from the news video. The utterance was "*hate is a motivation to do a crime,"* which was the opinion of a lawyer who featured in the video. Meanwhile, in Week 10, the participant used an utterance by the President of Indonesia in responding to a recent earthquake in the country. This was "*the situation is under control."*

In general, the content and composition of Student 20's speech performances changed qualitatively following the phase transition in her speech development. In the performances, the student was able to produce more content which was similar to the information presented in the news video materials. The composition of the reports' bodies also changed from the first to the second phase of the transition. The student was able to provide more detailed information and used more specific phrases from the videos in the second phase.

8.2.1.4 Conclusions regarding Student 20 from the explicit group

In conclusion, Student 20's speech fluency development experienced a clear phase transition and entered the attractor state. A sudden jump in Week 7 indicated this phase transition, after which Student 20 produced relatively longer speech with more syllables in her reports. The moving min-max graph and the Monte Carlo analysis confirm the significant difference in the development of the number of syllables in speeches before and after the phase transition. Critical fluctuation experienced by the moving standard deviation residuals of the number of syllables in the vicinity of the phase transition also confirms this finding.

Lastly, when the two phases of development were qualitatively compared, the content and composition of the speeches offered evidence of changes. Two speech fluency measures used for the content analysis, pause and repetition ratios, indicated substantial change from the first to the second phase of development. In the second phase, the participant produced a decrease in pausing, while increasing the repetition. The participant was able to produce more syllables in her speeches while at the same time reducing the number of pauses. In addition, the participant could generate more content, following the information presented in the news video materials. Meanwhile, the composition of the speech also indicated a change from the first to the second phase which is evident in the body of the composition and the use of borrowed utterances.

8.2.2 The results from Student18 (Explicit Group)

8.2.2.1 Sudden Jump

The change point analysis of Student 18's data showed a different outcome compared to Student 20. The speech rate, phonation time ratio, mean length of run, mean duration of pause, and number of syllables were analyzed by using the change point analyzer. However, no changes were detected in any of these variables. Therefore, further investigation by using the moving max-min graph was not applied for this student's data. However, in order to see any potential statistical differences in the data points, a Monte Carlo analysis was also applied. Consistent with the analysis of Student 20 data, all data points for Student 18 were also divided into two phases. Phase 1 was the period from Week 1 to Week 6, and

Phase 2 was the last three weeks of the data points. The analysis showed that there were no statistically significant differences between the two phases.

8.2.2.2 Anomalous Variance

Despite the undetected phase transitions on all the data points in Student 18's development, investigating the anomalous variances in the data would still be beneficial for this exploratory study of a complex dynamic system in speech recall and production. Anomalous variance is an indication of a fluctuation in a system's development. This fluctuation often signals developmental patterns, which could provide the basis to solve problems in the system (Kelso, 1995). Similarly to Student 20, to analyze the anomalous variance in this participant's data, the number of syllables was selected. Hence, a comparison could be made between the two participants. Figure 8.5 displays the moving standard deviation residuals of each moving window in Student 18's number of syllables per speech.

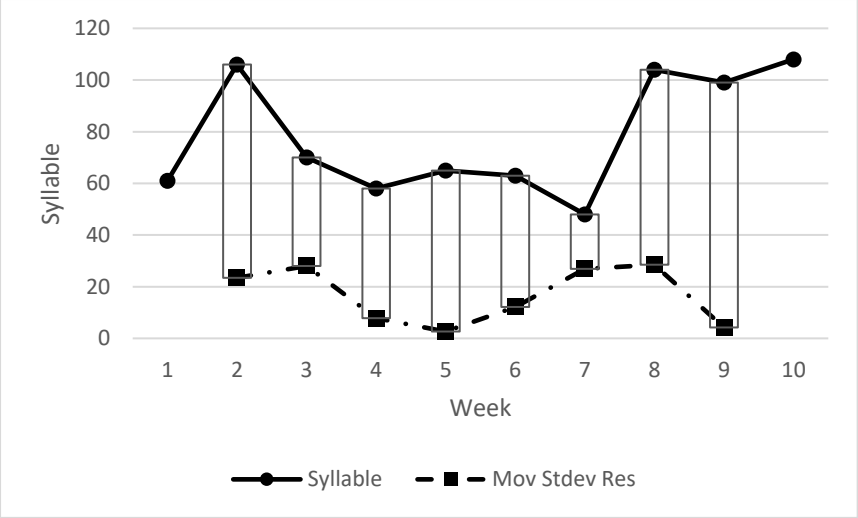


Figure 8.5 Magnitude of Change in the Number of Syllables of Student18

Note. The up/down bars represent the magnitude between the moving standard deviation residuals and their absolute values.

In contrast to Student 20's data, Student 18's moving standard deviation residuals fluctuated from the beginning to the end of the time series data. In other words, there was

no notable difference in the magnitude of the fluctuations, as found in Student 20's data. The smallest magnitude is seen in Week 7, while for Student 20, this occurred in Week 5. There was no indication of a critical fluctuation in Student 18's data, indicating no system preparation for a phase transition (sudden jump). This result was in line with the change point analysis, which resulted in no indication of sudden jumps.

8.2.2.3 Qualitative changes in the attractor

Even though Student 18's speech fluency development did not indicate a phase transition, the qualitative analysis was still conducted. Apart from being exploratory, the purpose of this analysis of a complex dynamic system was to provide data regarding systems development when no phase transitions occur. The data can be compared with those of Student 20 who experienced a phase transition. This comparison could provide some confirmation of the validity of attractor states, phase transitions, and the basis for understanding individual L2 learners' speech development in general. For this purpose, the analysis was similar to what was conducted for Student 20. The analysis comprised an investigation of changes in pause and repetition ratios, and changes in the speech composition and contents.

Changes in pause and repetition ratios

The change in pause and repetition ratios for Student 18 is detailed in Figure 8.6. The participant's pause ratio was at the highest (56%) when she started the practice in Week 1. This ratio decreased sharply in Week 2 (16%) and then a slight fluctuation followed from Week 3 to Week 4. A gradual upward trend from Week 5 to Week 7 can be seen before another fluctuation took place for the rest of the practices. Interestingly, although the pause ratio reached another peak at Week 7, this number was still lower than Week 1. Also, the ratio in Week 10 reached another trough in the data points, but the percentage was slightly higher than Week 2, which marked the lowest percentage in the data. In general, the polynomial trend line in pause ratio also shows a decreasing trend, indicating that the participant's pauses decreased over time. However, this result did not mean a development shift to a higher level because no actual phase transitions occurred in the data.

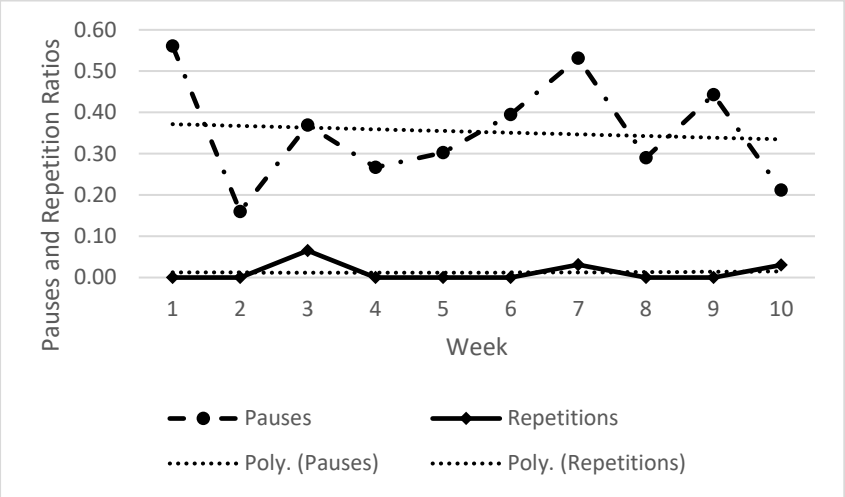


Figure 8.6 Ratios of pauses and repetitions against the number of syllables for Student 18 with the polynomial trend line (2nd degree)

Note. The ratios of pauses and repetitions are the result of dividing the variables by the number of syllables in the pruned transcriptions.

On the contrary, the participant's repetition ratios did not show much change during the ten data points. The participant only produced repetitions on three occasions: Week 3 (7% ratio), Week 7 (3% ratio), and Week 10 (3% ratio). It can be concluded that there was no substantial change in Student 18's repetition ratio during the treatment, indicating that this problem in her speech fluency of creating repetition during the speech persisted throughout.

Changes in contents and composition of the speech

Contents. Student 18's speech content looked similar from Week 1 to Week 7, in which mainly, the participant only reported one fourth to half of the information. The speeches showed some improvement in Weeks 8, 9, and 10, in which the participant was able to report up to two thirds of the information. In Week 1, for example, the participant only reported one segment of the information in the speech (see the following transcription). Meanwhile, the original news video provided five segments (see the discussion on Student 20 above regarding the original video's segments):

> *OK good morning ladies and gentlemen. Today, I will inform the news about terrorist.*

> *Some say, "no no I'm not terrorist". And then Islam is Muslim, and Islam not terrorist. What happen yesterday occur explosion and place crowded and then panic. (2) (Student 18)*

It can be seen from the transcription that after giving her introduction, the student only reported the second segment from the news video. This segment contained comments from several interviewees, yet the participant only reported the comment from one. Similar speech content was found from Week 2 to Week 5. The participant only conveyed one to two segments out of the five available on these two occasions. However, the participant's speech showed some improvement from Week 8 to Week 10. During this period, she consistently reported more than half of the provided information. In Week 8, the participant conveyed three out of five segments of information, the first, second, and fifth:

> *I will inform about the domestic violent in University of Memphis. (1)*
>
> *This incident happened on twenty thousand twelve. She was hit by her husband at their first year of marriage. They were fine but thing change for an instance. Her several tour for several year. She was very lucky because she was saved by his neighbour. (2)*
>
> *And now he was in the field of motivation about the domestic violent. (5) (Student 18)*

In Week 10, the participant also reported three out of five segments of information, as can be seen in the following transcription:

> *Ok good morning everyone. I'm Rina Puji Astuti from Queen TV. I will to inform about Indonesia earthquake. Tsunami warning triggers panic. This happen in Aceh, West Sumatera. The threat of tsunami hitting Aceh has triggered desperate panic. With many coastal resident evacuated their home and plan to higher ground. (1)*
>
> *The earthquake in this incident was felt until India and Thailand. (5)*
>
> *The resident trauma because of the incident at twenty thousand four and cause very strong panic. (4)*
>
> *Ok, I'm Rina Puji Astuti, back to studio. (Student 18)*

Here, the participant reported the first, the fifth, and the fourth segments consecutively. She did not mention the second segment regarding the US geological survey report about the earthquake, nor the third, which details the Indonesian President's press conference right after the earthquake. In addition, the participant only reported the summaries for the fourth and fifth segments, excluding other information from the original news report. The details regarding informational segments and the transcript of each video material can be found in Appendix 3.

Compositions. Student 18 did not show any changes in the composition of her speech through the ten weeks. The introduction, body, and conclusion of the composition remained the same. As part of the introduction, she usually mentioned her full name, the television station, and the news headline. On some occasions, she mentioned the date of the accident as well as the place. Similarly to Student 20, Student 18 did not follow the general guidelines for a news introduction. The following transcriptions are examples of Student 18's introductions:

Introduction in Week 4:

> *"Ok. Good morning people. I'm Rina from Queen News channel. I will inform about hail storm on the beach in America."*

Introduction in Week 8:

> *"Ok good morning everyone. I'm Rina Puji Astuti from Queen Channel. I will inform about the domestic violent in University of Memphis. This incident happened on twenty thousand twelve." (Student 18)*

As mentioned above, the body of the compositions in the participant's report contained less information than the news video. She only answered some of the guideline questions (5 W's and H) in presenting news (see Section 8.2.1.3). Sometimes the participant only reported the situation in the incident, including the victims. She rarely mentioned the cause and the time of the event. Following the same process as that used for Student 20, a comparison between Student 18's reports for Week 5 and Week 7 oshowed no differences in the body of these compositions:

Body in Week 5:

> *"*The first ever the fire that happen explosion. Nineteen killed all people feel panic. And national police provide assistance for this incident. A lot of safety coming down to help. With help of them resolved.*"*

Body in Week 7:

> *"*Cruise ship from Spain headed Portugal. They carry one thousand five hundred passenger. The cause big ship making and then three person felt injury.*" (Student 18)*

In reporting an airplane accident in Nepal in Week 5, in the body of the report, the participant did not mention what happened. Instead, she stated it briefly in the introduction. In the body, she only mentioned the victims (the 'who'). Similarly, in Week 7, with the report about a cruise ship accident, the participant did not mention what happened and how. She only mentioned where and who the victims were. In both reports, the participant did not include two crucial pieces of information: the 'what' and the 'how.' This occurred in most of the participant's speeches over the ten weeks.

In the conclusion, the participant did not produce any supplementary information in any of her reports. Mostly, she only closed with "*thank you*," repeated her name, and sometimes added an additional utterance, "*Back to studio."* The examples of the conclusions can be seen in the Week 4 and Week 10 transcriptions below:

Conclusion in Week 4:

> *"Okay, thank you. I'm Rina back to studio."*

Conclusion in Week 10:

> *"Ok, I'm Rina Puji Astuti, back to studio." (Student 18)*

In terms of borrowed utterances, the participant's speeches did not show any changes over the ten weeks. She only borrowed utterances from the video in three weeks: Week 1, Week 3, and Week 10. In Week 1, the participant used an utterance from one of the interviewees

in the news video. The interviewee said, "*no no no, I'm not terrorist….Muslim is not terrorist, and terrorist is not Muslim*". In her Week 1 report, the participant stated: *"no no no I'm not terrorist. Islam is Muslim, and Islam not terrorist."* In Week 3, the specific utterance borrowed by the participant from the news video was the same as Student 20's: "*a truck lost control."*

In her report for Week 10, the utterance borrowed by the participant was "*tsunami warning triggers panic*" which was also the additional headline written at the beginning of the video. However, this evidence of borrowed utterances was not considered as a developmental change since they occurred from the beginning (Week 1) to the end (Week 10) of the experiment, with no evidence in the middle.

8.2.2.4 Conclusions regarding Student 18 from the explicit group

To summarise, Student 18's speech development did not show any indication of change during the ten weeks of instruction. The content and composition of her news reporting speech performances did not undergo any qualitative changes throughout the study. The participant did not improve the way she paused during the speech. Most of her speeches also had a similar amount of repetition. This finding corroborated the result of the change point analysis in which there was no sudden jump detected in the participant's time series data.

8.2.3 The results from Student8 (Implicit Group)

8.2.3.1 Sudden Jump

Like participants in the EG, the same five speech fluency measures were also analyzed to investigate sudden jumps in the data for participants in the IG. The change point analysis of Student 8 showed that only the mean duration of pauses (MDP) changed. The MDP was calculated by dividing the total length of pause duration with the total pause number (see Section 6.6.1 for the details). This data point experienced one change within the time series. As detailed in Figure 8.7 and Table 8.3, the change occurred in Week 10 with a 93% confidence level. The confidence interval was between Week 7 and Week 10, which encompasses when the actual change occurs. Before the change, the average value of the

MDP was 0.59. After the change, the average became 0.22. This was classified as a Level 3 change because it was detected in a third pass through the data after the data were subdivided into three subsets in the analysis. This level indicated that the change was relatively small.

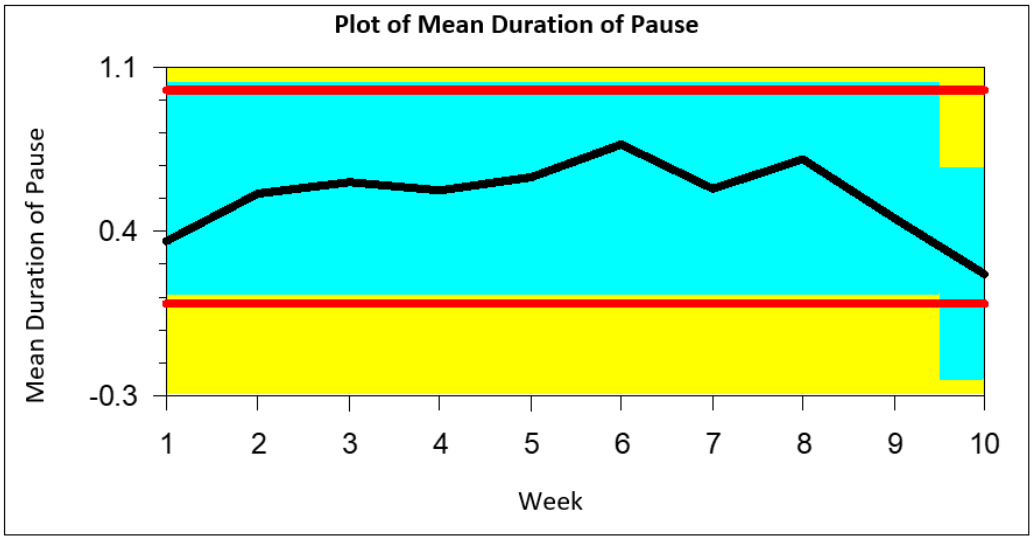


Figure 8.7 Change point analyser plot of number of syllables produced by Student 8

Note. 1. The red lines are control limits representing the maximum range over which the values are expected to vary, assuming no change has occurred.
2. Shifts in the blue region represent changes in the time series data.

Table 8.3 Change point analysis result for mean duration of pauses of Student 8

Detected Change	Confidence Interval	Confidence Level	Average Change of Number of Syllables		Level of Change
Week	**Week**		**From**	**To**	
10	(7, 10)	93%	0.59	0.22	3

Note. 1. The mean duration of pauses is the result of dividing the total length of pause durations by the total pause number.
2. The average change is based on the average number of syllables in each phase (the first phase consists of Week 1 to Week 9, and the second phase comprises Week 10 only).

Further analysis by using the moving min-max graph for the MDP is presented in Figure 8.8. As can be seen in the figure, the MDP development was nonlinear. It seemed to increase steadily from Week 1 to Week 6 before fluctuating for the next two weeks (Weeks 7 and 8). After Week 8, the value plummeted, reaching its lowest point in Week 10. As confirmed by the change point analysis result, the MDP value in Week 10 (0.22) was lower than the average value of 0.59. This value was also lower than the value in Week 1 (0.36), which represented the initial condition of the participant's MDP.

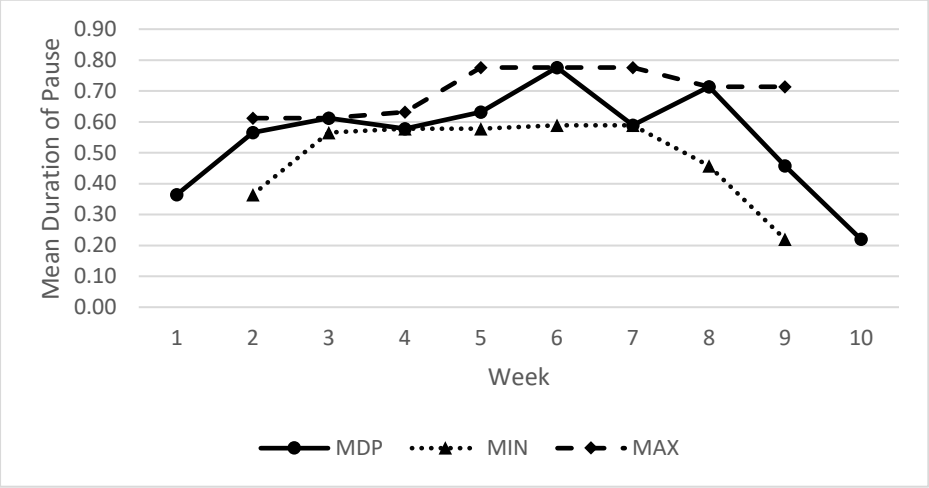


Figure 8.8 Change in mean duration of pauses for Student 8 shown by using moving min-max graph

Note. The mean duration of pauses is the result of dividing the total length of pause durations by the total pause number.

The participant's MDP trajectory also indicated periods of progress and regression, as represented by the fluctuations in the time series data. Although the change point was detected at the end of the period, the line graph shows that the MDP trajectory noticeably declined after Week 8. The bandwidth between the MDP value and its minimum and maximum values enlarged after this week, with Week 9 manifesting the largest bandwidth. The fluctuations from Week 6 to Week 9 and the large bandwidth in Week 9 could be considered as the pre-conditions of a phase transition. Hence, this result of the moving min-max graph confirmed the change point analysis, which stated that a significant change in the time series data occurred in Week 10. However, a further analysis by using Monte Carlo

analysis could not be applied to this data because the change point happened at the end of the series.

8.2.3.2 Anomalous Variance

Figure 8.9 shows the moving standard deviation residuals of each moving window of Student 8's MDP development. There was no noticeable fluctuation in the magnitude of moving standard deviation residuals that could indicate the system's preparation for a phase transition. The highest standard deviation was detected in Week 9, in which the value (0.20) was close to the minimum value (0.22). The moving standard deviation residuals shown in Week 9 were also the highest in the data series. Therefore, the anomalous variance analysis on Student 8's MDP did not support the change point analysis result. This inconclusive result could be caused by the very late sudden jump (Week 10), as detected by the change point analyzer, which only occurred at Level 3.

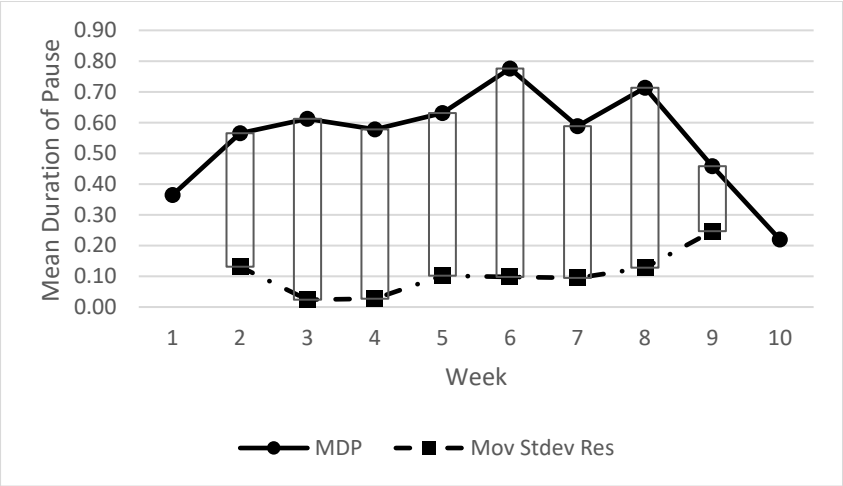


Figure 8.9 Magnitude of change in Student 8's mean duration of pauses

Note. 1. The mean duration of pauses is the result of dividing the total length of pause durations by the total pause number.

8.2.3.3 Qualitative changes in the attractor

Since the change point was detected at the end of the time series data (Week 10), no data points could be analyzed to describe the development's condition in the attractor (after the sudden jump). Instead, the data points before the sudden jump were selected for analysis purposes. This decision was made in order to collect evidence regarding the system preparation to enter a phase transition. Also, the analysis would be useful in revealing the participant's internal development. Initial analysis based on reading all the transcripts of participant's reports showed visible differences in the learner's speech samples from Week 8 to 10. These three weeks were compared with the seven previous weeks. In the analysis, the data points from Week 1 to Week 7 were named the first cluster (instead of phase), while data points from Week 8 to Week 10 represented the second cluster.

Changes in pause and repetition ratios

Figure 8.10 illustrates the change in pause and repetition ratios in Student 8's speech samples. Similarly to Student 18 from the EG, Student 8's highest pause ratio (33%) occurred in Week 1. The ratio decreased sharply in Week 2 to 9%, the lowest ratio in the data point. This was then followed by continuous fluctuations from Week 3 to Week 10. The trend line shows a decrease in the participant's pause ratio. However, this trend line does not indicate changes in the participant's pause ratio from the first to the second cluster since the trajectory looks similar from the beginning to the end.

Meanwhile, the ratio of repetition did not demonstrate much change during the data series. The participant created a 2% ratio at the beginning (Week 1) and ended up with 1% at the end (Week 10). In general, the ratios fell within a small range of 0 to 2% because the participant did not create many repetitions in the speeches. In fact, there was no repetition in Weeks 2, 3, 7, 8, and 9.

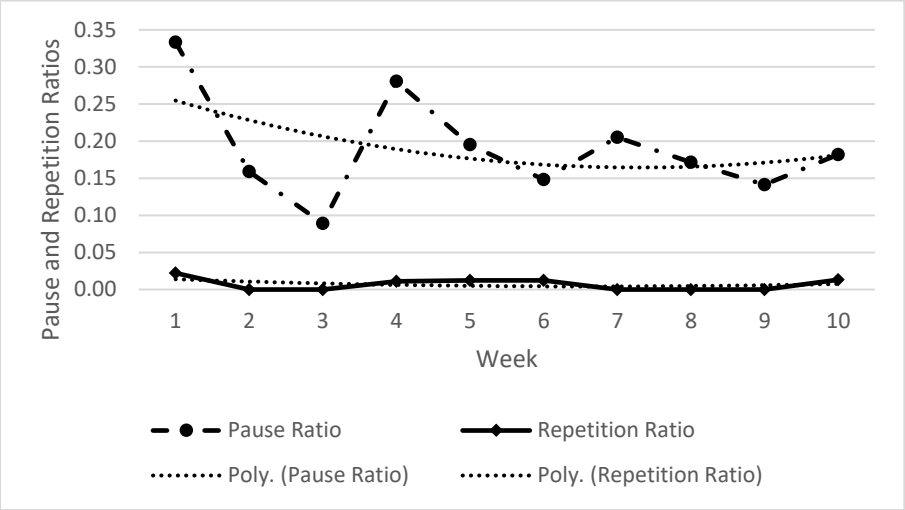


Figure 8.10 Ratios of pauses and repetitions against the number of syllables for Student 8 with a polynomial trend line (2nd degree)

Note. The ratios of pauses and repetitions are the result of dividing the variables by the number of syllables in the pruned transcriptions.

Changes in contents and composition of the speeches

Contents. In terms of content similarity to the actual news video, Student 8's speech was more elaborate in the second cluster where the speech samples included more detailed information than in the first cluster. The longest speech produced by the participant was found in Week 8 in which the news video was about domestic violence (see Appendix 3 for the transcript). The news consisted of five segments of information. The first segment could be considered as the headline of the news. Details regarding what happened to the victim of the domestic violence were given in the second segment. The third segment was a report about how the victim got out of her situation. The fourth segment comprised opinions of the attendees towards domestic violence cases. In the last segment, the victim's message for others was quoted about getting out of any domestically violent situation. As can be seen in the following transcription, the participant reported all segments:

> *Good morning people. This is Zikra Darni will report to you about domestic violence survivor.*
>
> *A former Memphis news anchor who survived domestic violence is sharing her story. (1)*

She didn't think domestic violence could happen to her until what a start as a fairy tale marriage in two thousand and ten and with her fighting for her life. It happen when her baby was born in May of two thousand and eleven but then during of the fifth week, when their son was five weeks old, he snapped. Her ex-husband snapped. And he physically abused her, hands around her neck. She said, a neighbour who heard her screaming called a police. She said her husband pinning her up against the refrigerator and squizzing her neck were she couldn't breathe. (2)

And then she made it out of her marriage and said her son gave her strength to live. She said that this could happen to anyone. (3)

She was a professional who had everything going on in life and it happen to her. (4)

She said get out, save yourself, save your family. (5)

That's all my report. Thank you. (Student 8)

Similarly to Student 20 from the explicitly instructed group, the participant reported all the four segments of information in her speech in Week 9 (see transcription below). The news video was about hate crimes in Canada (see the relevant discussion on Student 20 above). Student 20's report on this news was more detailed than Student 8's, who recounted only the summary of each segment:

Good morning people. I will report to you about the case of hate crime.

The hate crime incident happening outside Walmart near Hamilton over a parking space. Begin with the man on the left confronting the driver. The man's wife is taking the video appears to get hit by the truck. 1

The police said, they are reviewing it as possible hate crime. 2

And the case of hate crime also happen in Toronto. Toronto police say they do consider this a hate crime. 3

The video show a debate on social media among the issue, were they hate crime, and if so what would the consequences be a legal expert. 4

That's all my report. Thank you. (Student 8)

In Week 10, however, the participant only reported on three out of five segments of information in her speech (see transcription below). The news this week was about an earthquake in Aceh, Indonesia, which consisted of five segments of information. The first segment informed the audience about the situation in the city of Aceh right after the

earthquake happened. The second segment contained information from the United States Geological Survey (USGS) regarding the earthquake's strength and epicentre. The third segment reported a press conference from the President of Indonesia updating the latest condition in Aceh, several hours after the earthquake. In the fourth segment, the news mentioned a deadly earthquake which also happened in Aceh in 2004. Finally, the last segment informed about the automatic activation of the tsunami early warning system around the Pacific Rim right after the earthquake.

> *Good morning people. I'm Zikra Darni will report to you about Indonesia earthquake. The earthquake strike Aceh and people get panic about the tsunami issue. And they evacuate to several place and high place. (1)*
>
> *The United States Geological Survey said that the earthquake has eight point six magnitude. (2)*
>
> *There is no victim right now and some people got injured. The President of Indonesia Susilo Bambang Yudhoyono said that the situation is under control and the people in Aceh get to safety place. (3) (Student 8)*

On the other hand, the participant's speech samples in the first cluster contained less detailed information. In Week 1, for example, she only reported information from two out of the five segments in her speech (see the discussion on Student 20 above for details regarding the news). These were the first and third segments:

> *Good morning people. This is Zikra Darni reporting from the location. You know that Indonesia is a country that always get terror. (1)*
>
> *And we also know that Indonesia is the Muslim country.*
>
> *And as the narasumber* [source] *said, I rejected the terrorist in the name of Islam. Moslem is not terrorist, and the terrorist is not Muslim. (3)*
>
> *Thank you. (Student 8)*

In the following weeks, the extent of the participant's content fluctuated from Week 2 to Week 7. After reporting only two out of five segments in Week 2, she began to convey more information (three out of five segments) in Weeks 3, 4, 5, and 6. In Weeks 4 and 5, her speeches also included more details. However, the content dropped again in Week 7, in which the participant only reported two segments out of four.

Compositions. Student 8's compositions demonstrated a substantial change in the second cluster in the body and the specific utterances borrowed from the news items. However, for the introductions and conclusions, a similar pattern can be observed throughout. In the introduction, the participant usually introduced herself by mentioning her full name and adding the phrase "reporting from the location." This method of introduction occurred in Weeks 1 through 4 and again in Week 6. In Week 5, however, she added a short headline to the introduction, which also occurred in Week 7 to Week 10. The examples of the introduction can be seen from the participant's performances in Week 1, Week 5, and Week 10 below:

Introduction in Week 1:

"Good morning people. This is Zikra Darni reporting from the location."

Introduction in Week 5:

"*Good morning everyone. Zikra Darni will report about Nepal airline crashed*."

Introduction in Week 10:

"Good morning people. I'm Zikra Darni will report to you about Indonesia earthquake." (Student 8)

Despite the additional short headlines, the participant's introduction still did not follow the general guidelines of a news introduction (see Section 8.2 regarding the general guidelines).

In her speech's conclusions, the participant signed off but did not follow the suggested guideline of including supplementary information. In most of the performances, she said, "*thank you*" with some additional phrases "*This Zikra Darni reporting from the location,"* "*Back to the studio,"* and "*That's all my report.*" A consistent conclusion could be found in Weeks 7 to 10. Here, the participant gave the same formulaic (and slightly inaccurate) conclusion, "*That's all my report; thank you*" at the end of each speech, with no supplementary information.

Meanwhile, the body of Student 8's compositions showed some changes in the second cluster. Unlike the first cluster, complete information regarding answering the 5 W's and H questions in detail could be found in the second phase, such as in Weeks 8, 9, and 10. A

comparison of the participant's performances between Week 2 (terror incident) and Week 8 (domestic violence) helps highlight this change in the body of the speech composition. In Week 2, her report regarding a terror incident in France lacked details (see the transcription below). There was no information regarding the actual process of the incident. The participant only reported general information about the victims, the suspected terrorist, and the place the incident occurred. She did not specify what, when, and how the event happened:

> *There is a nasty condition in in France. The gunman claiming link to Islam in southern France. This bombing in supermarket southern in France. That bomber take hostage in France. The tragedy killing one people, one police office and two people between the supermarkets. (Student 8)*

In Week 8, on the contrary, the report was full of details regarding the actual incident. In reporting news about domestic violence, the participant conveyed detailed information from the beginning to the end, following the sequence of information provided in the news video. It began with what happened to the victim, when the violence occurred, how it happened, and how it stopped. The participant closed her report with a quote from the victim. This was the same closing used in the news video:

> *A former Memphis news anchor who survived domestic violence is sharing her story. She didn't think domestic violence could happen to her until what a start as a fairy tale marriage in two thousand and ten end with her fighting for her life. It happen when her baby was born in May of two thousand and eleven but then during of the fifth week, when their son was five weeks old, he snapped. Her ex-husband snapped. And he physically abused her, hands around her neck. She said, a neighbour who heard her screaming called a police. She said her husband pinning her up against the refrigerator and squizzing her neck were she couldn't breathe. And then she made it out of her marriage and said her son gave her strength to live. She said that this could happen to anyone. She was a professional who had everything going on in life and it happen to her. She said get out, save yourself, save your family. (Student 8)*

As can be seen from the transcription above, the way the participant conveyed the speech was really similar to the news video itself (see Section 8.2.1.3). She detailed what happened to the woman regarding the domestic violence she experienced. She began by mentioning the woman as a violence survivor. Then, she continued recounting about how and when the

domestic violence started. She then followed by mentioning how it finished, including what happened during the violence. In the end, she quoted the woman's message for audiences to stand against this violence. Interestingly, Student 8's report was quite similar to Student 20's in terms of the detail in the information presented and the flow of the speech, which resembled the actual news video. This similarity might be influenced by the news content, which involved a woman as the victim, which is the gender of the participants.

Student 8 produced another set of detailed information in the body of her composition in Week 9's speech, covering most of the information presented in the news item about hate crimes in Canada (see Appendix 4 for Student 8's transcriptions). As in the news video, she began by reporting about the first incident involving a couple (husband and wife) with a man in a truck. She summarized the incident in her report and then talked about the local authority's action. She then continued by reporting about the second incident in another city and commented on social media reactions regarding the two incidents. The comments she added were not from the news video. This form of the body was again found in her report in Week 10 (see Appendix 3).

In terms of borrowed phrases, Student 8's speech showed significant differences when the two clusters of data points were compared. In the first cluster, she used some phrases from the news video in five weeks: Weeks 1, 2, 3, 4, and 7. In Week 1, the participant repeated two utterances from the interviewees in the video. In one utterance, the participant said, *"I rejected the terrorist in the name of Islam,"* repeating the original utterance from one of the interviewees who said, *"I reject this terrorism in the name of Islam."* The second utterance repeated by the participant was, "*Muslim is not terrorist, and the terrorist is not Muslim."* This utterance was originally, *"Muslim not terrorist, or terrorist not Muslim."* In Week 2, the participant only borrowed one utterance from the news video. She mentioned the utterance *"nasty condition,"* which was a rephrasing of the utterance *"very nasty situation"* in the video. In Week 3, she borrowed two utterances from the video, stating "*tragic way"* and *"tyre blow."* The actual utterances were "*tragic day"* and "*tyre blow up."* The participant borrowed one utterance each in Week 4 ("*freezing temperature"*) and Week 7 (*"safety*

drill"). Meanwhile, no borrowed utterances were found in the participant's speech in Weeks 5 and 6.

In Week 8, when the second cluster started, the participant produced a much more similar speech to the news video, borrowing eight utterances, as can be seen in the following transcription:

> *She didn't think domestic violence could happen to her until what a start as a fairy tale marriage in two thousand and ten end with her fighting for her life*
>
> *During of the fifth week, when their son was five weeks old, he snapped. Her ex-husband snapped.*
>
> *He physically abused her, hands around her neck.*
>
> *A neighbour who heard her screaming called a police.*
>
> *She made it out of her marriage and said her son gave her strength to live.*
>
> *This could happen to anyone. She was a professional who had everything going on in life and it happen to her.*
>
> *Get out! Save yourself! Save your family! (Student 8)*

Interestingly, these utterances were the exact ones used in the video. The participant just changed the first-person narrative "*I"* and *"my"* to the third person "*she"* and "*her"* in the speech. Again, this kind of speech production was only found in a speech sample reporting domestic violence against a woman.

In Week 9, the participant borrowed five exact utterances from the news video:

> *Incident happening outside Walmart near Hamilton over a parking space.*
>
> *Begin with the man on the left confronting the driver.*
>
> *The man's wife is taking the video appears to get hit by the truck.*
>
> *The police said, they are reviewing it as possible hate crime.*
>
> *Toronto police say they do consider this a hate crime. (Student 8)*

Similarly to what happened to the content of the speech, the participant's borrowed utterances were reduced in Week 10, only borrowed three: "*the earthquake strike," "the situation is under control,"* and *"the people in Aceh get to safety place."* The first was actually borrowed from the first utterance in the video that began with "*As the earthquake*

strike…." The other two utterances were initially mentioned in the video by the President of Indonesia, who stated that *"The situation is under control. There was a bit of panic, but the people in Aceh have been moved to safety."*

8.2.3.4 Conclusion regarding Student 8 from the implicit group

Despite the detected sudden jump in Week 10, other analyses involving anomalous variance and qualitative change in the attractor did not find evinces of change. In general, the content and composition of her reporting speech performances did not undergo any qualitative changes throughout the study. The participant's pause and repetition ratios did not change from the first to the second cluster. The content and composition of her speech samples also looked similar in the two clusters. The only improvement found in the second cluster was in the form of a detailed body of the composition. However, other analyses were not supporting the change-point analysis result. Therefore, the sudden jump detected by the change-point analyzer was considered inconclusive.

8.2.4 The results from Student22 (Implicit Group)

8.2.4.1 Sudden Jump

The results of the change point analysis for Student 22 showed a sudden jump in the number of syllables. As can be seen in Figure 8.11 and Table 8.4, the change occurred in Week 8, with a 92% confidence level. The confidence interval was between Week 5 and 9 and this interval covered the actual change point at Week 8. Hence, the development of number of syllables could be divided into two phases. The first phase consisted of Week 1 to Week 7, and the second phase comprised Weeks 8, 9, and 10. The number of syllables produced by the participant before the change was 78.57 on average, which increased to 111 after the change. This occurred at Level 3, indicating a relatively small change.

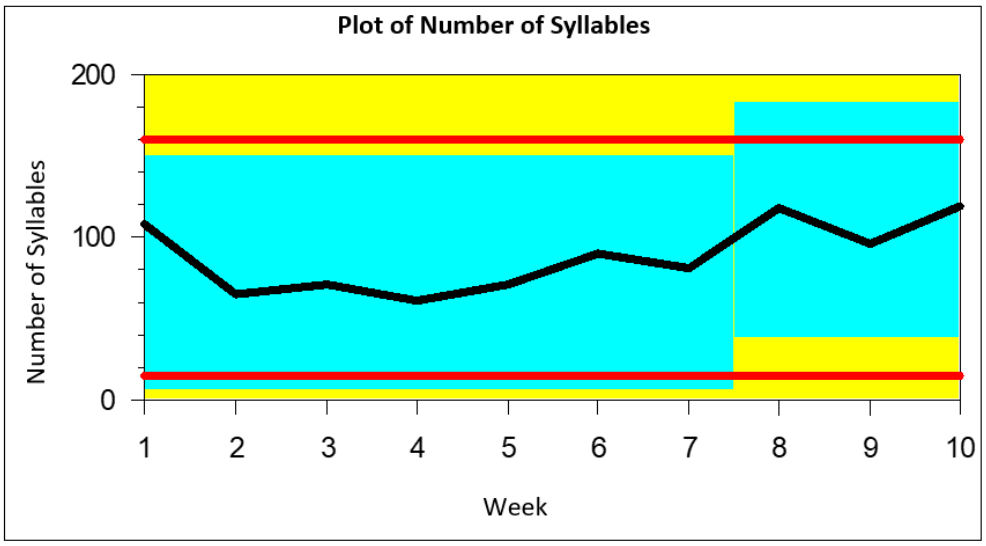


Figure 8.11 Change point analyser plot of number of syllables for Student 22

Note. 1. The red lines are control limits representing the maximum range over which the values are expected to vary, assuming no change has occurred.

2. Shifts in the blue region represent changes in the time series data.

Table 8.4 Change point analysis result for number of syllables of Student 22

Detected Change	Confidence Interval	Confidence Level	Average Change of Number of Syllables		Level of Change
Week	Week		From	To	
8	(5, 9)	92%	78.57	111	3

Notes. 1. The number of syllables is the actual number of syllables produced in the pruned transcription. 2. The average change is based on the average number of syllables in each phase (the first phase consists of Week 1 to Week 7, and the second phase is from Week 8 to Week 10).

Further analysis by using the moving min-max graph for the number of syllables, as presented in Figure 8.12, showed that the development of the number of syllables was nonlinear. This number plummeted from Week 1 to Week 2 before a minor fluctuation from Week 2 to Week 4. The number of syllables in Week 4 was the lowest (62) and fell below the average value of the first phase (78.57). After that, the number of syllables increased steadily up to Week 6 before a series of more significant fluctuations occurred from Week 7

to Week 10. In Week 8 the second highest number of syllables (118) was recorded, with Week 10 the highest (119). Values in these two weeks were also higher than the average value of the second phase (111), as indicated by the change point analysis.

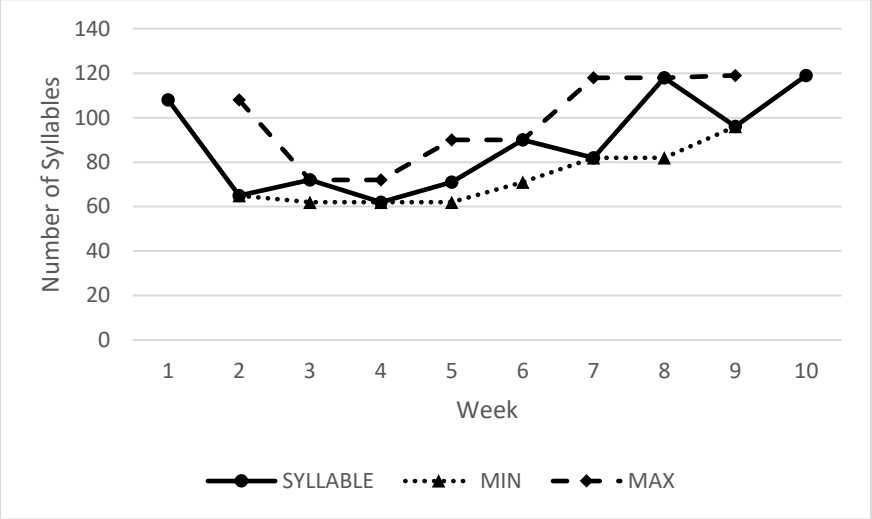


Figure 8.12 Change in the mean number of syllables for Student 22 shown by using moving min-max graph

As Figure 8.12 indicates, the bandwidth between the number of syllables and the minimum value enlarged at Week 8, when the largest bandwidth in the data was recorded. The fluctuations were even more prominent in this second phase. Despite the drop at Week 9, the general trend of the trajectory was rising.

Monte Carlo analysis showed a statistically significant difference between the two phases of development in terms of Student 22's number of syllables (p = 0.033). This result corroborated the change point analysis finding.

8.2.4.2 Anomalous Variance

Figure 8.13 shows the moving standard deviation residuals of each moving window of Student 22's development in the number of syllables enunciated in a speech. No substantial fluctuation was seen in the moving standard deviation residuals. As shown by the up/down bar, the smallest magnitude occurred at the beginning of the time series (Week 2). After that, the magnitudes increased steadily, with minor fluctuations to the end of the series.

The absence of high fluctuation in the magnitude of moving standard deviation residuals and the original values being in the vicinity of a change point indicates that the system did not completely shift to another phase, as suggested by the change point analysis. This result was understandable due to the fact that the change was only detected at Level 3 (see Section 8.2.4.1).

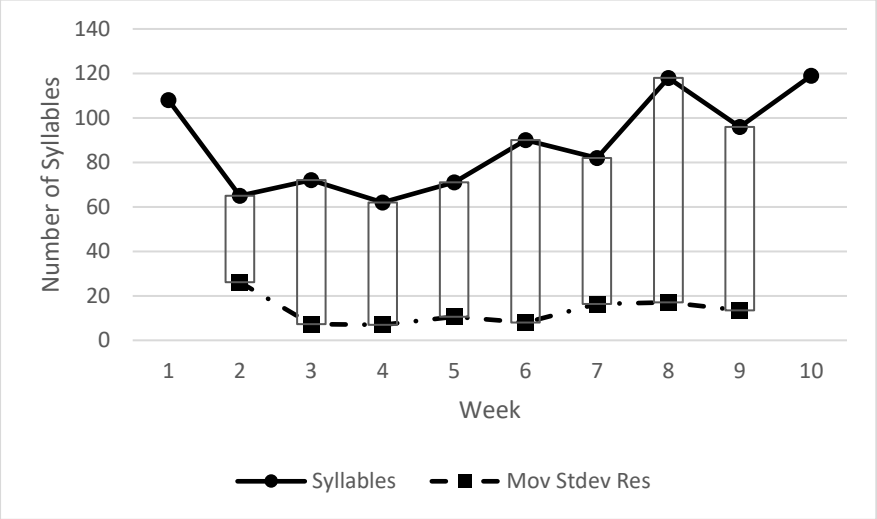


Figure 8.13 Magnitude of change for Student 22's number of syllables

Note. The up/down bars represent the magnitude between the moving standard deviation residuals and their absolute values.

8.2.4.3 Qualitative changes in the attractor

Despite the inconclusive result suggested by the anomalous variance analysis, a qualitative analysis based on changes in the attractor was conducted. The analysis involved a comparison of the two phases of development indicated by the change point analysis: Week 1 to Week 7 data for the first and Weeks 8, 9, and 10 for the second.

Changes in pause and repetition ratios

Figure 8.14 displays the change in pause and repetition ratios of Student22's speech samples. A downward trend in the pause ratios is indicated, with some fluctuations. The trend line also confirms this indication. The participant's pause ratios started at 30% and were 17% at the end. There was one peak in each of the phases, and the peak in the second

phase (pause ratio = 28%) was much lower than that of the first phase (pause ratio = 40%). Figure 8.14 also shows that the first phase consisted of two troughs that occurred in Week 3 (pause ratio = 19%) and Week 6 (pause ratio = 16%), which were also higher than the only trough found in Week 9 (pause ratio = 13%) in the second phase. The trajectory of pause ratios shows two different patterns in the two phases, indicating a change in the development, as suggested by the change point analysis.

On the other hand, there was no notable change in repetition ratios during the data series. Similarly to Student 8, Student 22 created a 2% repetition ratio at the beginning (Week 1) and finished with 1% at the end (Week 10). The only high percentage of repetition made by this participant was in Week 4 (5%), which was part of the first phase. Therefore, the way she repeated words or phrases in the speech remained the same for the entire time series data.

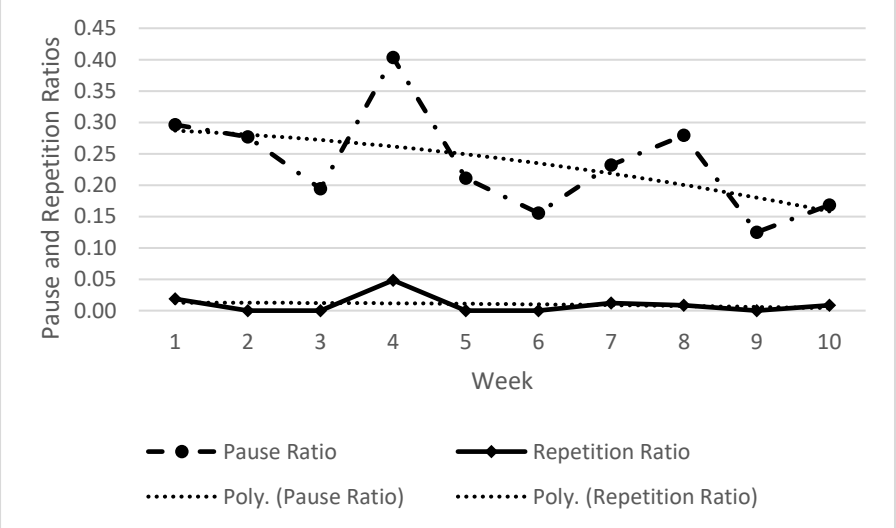


Figure 8.14 Ratios of pauses and repetitions against the number of syllables for Student 22 with the polynomial trend line (2nd degree)

Note. The ratios of pauses and repetitions are the result of dividing the variables by the number of syllables in the pruned transcriptions.

Changes in contents and compositions of the speech

Contents. The contents of the participant's speech samples looked similar across the time series. Mostly, the participant only reported two segments of information in her speech. The only exception was in Week 10, in which she reported four segments out of the five presented in the news. In reporting the news about an earthquake in Aceh, Indonesia (see Section 8.2.3.3 for the description about the news), the participant reported the first, third, fourth, and fifth segments with a reversed order between the last two. The only segment she missed was about the USGS report regarding the earthquake's magnitude and epicentre:

> *Good morning people, I'm Sandra Handayani Putri, will present the news for today. Indonesian earthquake in Banda Aceh. It is a big disaster. (1)*
>
> *Indonesian president Susilo Bambang Yudhoyono said that situation is under control. The nation so panic that very under control but it has make sure the sitting well. Government has to manage the disaster. (3)*
>
> *This earthquake also make people in Thailand and India afraid. They expect that Thailand and India will flow by tsunami. (5)*
>
> *And people remember tsunami in two thousand and four. (4)*
>
> *That's the news for today. Thank you. (Student 22)*

As a comparison, the participant's speech samples from Week 1 and Week 8, in which only two segments were reported, looked similar, although both weeks belonged to different developmental phases. In reporting a terror incident in Jakarta in Week 1, the following transcription shows how the participant only mentioned the two information segments:

> *Good morning audience. To the news about Jakarta terror.*
>
> *The accident is so large and the Islam is something be caught because people call it is very different from Islam. (1)*
>
> *All the people said that it's from Islamic centre. People say that Muslim in Indonesia isn't terrorist and terrorist is not Islam. And in the twenty fourteen there is happen the same accident or the same problem and it Islamic state. So people judge Islamic community is the terrorist. (2)*
>
> *Thank you (Student 22)*

Similarly, the participant also reported only two segments in Week 8:

> *Good morning people. I'm Sandra Handayani Putri, will present news about domestic violence.*
>
> *The woman said that her husband was very good at first before marriage. But after five months having a baby, her husband snapped and turning into a violent one. The woman usually hidden and scream, but fortunately saved by the neighbour from the violence in here in her household. (2)*
>
> *So, she can save herself since that. (3)*
>
> *Actually in two thousand eleven the incident have made people felt delay a lot of time.*
>
> *That's all news today. Thank you. (Student 22)*

Therefore, this indicates that no notable change could be seen in terms of the contents of the speeches.

Composition. The compositions of Student 22's reports showed a different pattern of introduction in the second phase, while the body and conclusion remained similar. There was also a slight difference in how the participant used borrowed utterances in the second phase. Student 22 only made a greeting and mentioned the news headlines in her introduction in the first phase. In the second phase, she began to introduce herself by mentioning her full name in Week 8, 9, and 10. The following examples of introductions from Week 1, 5, 8, and 10 illustrate this difference:

Introduction in Week 1:

*"*Good morning audience. To the news about Jakarta terror."

Introduction in Week 5:

"*Good morning audience. We come with the news."*

Introduction in Week 8:

"Good morning people. I'm Sandra Handayani Putri, will present news about domestic violence."

Introduction in Week 10:

"Good morning people, I'm Sandra Handayani Putri will present the news for today. Indonesia earthquake in Banda Aceh" (Student 22)

In the conclusions, the participant did not include any supplementary information. The speeches during the first phase were closed mainly with "*thank you*." In the second phase, starting from Week 8, the participant gave a similar concluding phrase preceded by an additional utterance, "*That's all for today. Thank you".* This kind of conclusion was repeated in Week 9 and Week 10.

Meanwhile, the body of the compositions became complete only in Week 8. However, this dropped again in Week 9 and Week 10, which looked similar to the compositions in the first phase. The participant answered most of the 5 W's and H questions in her speech in Week 8 by mentioning what and when the incident happened, why it happened, who was involved, and how it ended.

> *The woman said that her husband was very good at first before marriage. But after five months having a baby, her husband snapped and turning into a violent one. The woman usually hidden and scream, but fortunately saved by the neighbour from the violence in her household. So, she can save herself since that. Actually in two thousand eleven the incident have made people felt delay a lot of time. (Student 22)*

In Week 9, similar detail was not found in the body. The participant only reported half of the news and added her own comments. As stated above, in analyses of other students' speeches, Week 9's news was about two hate crimes in two different cities in Canada. However, the participant only reported one case:

> *It is in Toronto, Canada. That is near the hospital one of possible hate crime. He didn't expect this while taking the video that viral. The video was upload in social media, so people know. In this era, everything you do involving picture or video in public are seen. And it cannot hidden by violence alone. (Student 22)*

Although slightly better than Week 9, the report in Week 10 lacked detail in terms of covering the 5 W's and H questions. This news item was about an earthquake in Aceh, Indonesia (see Section 8.2.3.3). However, no information was recounted regarding what happened in the city right after the earthquake. Moreover, Student 22 did not include the

report from the USGS, which was considered vital information regarding an earthquake disaster.

> *Indonesian earthquake in Banda Aceh. It is a big disaster. Indonesian president Susilo Bambang Yudhoyono said that situation is under control. The nation so panic that very under control but it has make sure the sitting well. Government has to manage the disaster. This earthquake also make people in Thailand and India afraid. They expect that Thailand and India will flow by tsunami. And people remember tsunami in two thousand and four. (Student 22)*

The way the participant reported the body of the compositions in Week 9 and 10 was similar to the first phase. In Week 2, for example, her report was concise. Despite mentioning the suspect and the location of the terror incident, she gave no information regarding how exactly the incident happened, mussing much of the details:

> *Suspected ISIS terrorist takes in a supermarket in the Southern. The gunman claimed holding hostage in France. One person there killed near the supermarket and shot dead two other people as appear. (Student 22)*

There were no changes in terms of the borrowed utterances in Student 22's reports after the phase transition. In general, the participants borrowed utterances in six weeks, which spanned part of the two phases of development. Using one each week, the participant borrowed utterances in Week 1, 3, 4, 5 in the first phase; and in Week 8 and 10 in the second phase. In Week 1, she stated, "*Muslim in Indonesia is not terrorist and terrorist is not Islam.*" The original utterance in the video was, "M*uslim not terrorist or terrorist not Muslim."* In Week 3, she said, "*the truck have a tyre blow*," a similar utterance to that used by Student 8. In Week 4's video, the news included an utterance, "*then suddenly everything change*," and the participant used only the last part in her report by saying, "*everything change."* In Weeks 6, 8, and 10, the participant used one exact utterance from the videos in each of the reports. She mentioned "*the smoke bellow into the sky"* in Week 6, "*her husband snapped"* in Week 8, and "*situation is under control"* in Week 10. The similarity found between the two phases proved that there was no change in Student 22's development in terms of borrowing utterances as she only used one per week each time.

8.2.4.4 Conclusion regarding Student 22 from the implicit group

Student 22's speech fluency, as represented by the number of syllables, showed a developmental change indicated by a sudden jump in Week 8 of the time series data. This finding was supported by the moving min-max graph that showed high fluctuations between the number of syllables and its min-max values. The follow-up Monte Carlo analysis also corroborated the change point analysis finding. However, the anomalous variance analysis indicated that a phase shift in the system did not take place completely.

The results of further analysis involving a qualitative change in the attractor also varied. Analysis of the pause ratios indicated a change in the development over the two phases, as suggested by the change point analysis. However, the repetition ratios did not show any changes. In addition, the contents of the participant's speech samples looked similar across the entire instruction period. Meanwhile, the compositions of the reports only showed a difference in terms of their introduction, with the body and conclusion remaining the same, including the number of borrowed utterances used in each report. Therefore, the result of the analysis for Student 22's data was also inconclusive.

8.3 Summary of the Phase Transition Analysis Results

The results of the phase transition analysis involving the four participants with high gains from the two study groups were varied. The analysis was based on five speech fluency measures: (1) the speech rate, (2) phonation time ratio, (3) mean length of runs, (4) mean duration of pauses, and (5) number of syllables. It seemed that number of syllables was a more reliable predictor for phase transition analysis because the change point analyzer was able to detect a sudden jump for two participants, one from each group of the study. The development of the number of syllables for Student 20 from the explicit instruction group showed a sudden jump in Week 7 of the time series data. This finding was supported by the anomalous variance and a qualitative change in the attractor analysis. On the other hand, for the other participant from the same group, Student 18, no indication of a phase transition appeared in any of the five measures.

In the implicit instruction group, for one participant (Student 22) there was also an indication of a phase transition in the development of the number of syllables. The sudden jump experienced by this participant was detected in Week 8. However, this finding was not supported by the anomalous variance analysis. Yet, it was partly supported by the qualitative change in the attractor analysis, especially in terms of pause ratios and the introductions to the speeches. Meanwhile, for the other participant in this group (Student 8) a sudden jump in the development of the mean duration of pauses was indicated. However, since this was detected at the end of the time series (Week 10), follow-up analysis involving anomalous variance and qualitative change in the attractor could not be conducted. The supplementary analysis by grouping the data into two clusters also failed to detect any qualitative changes in the speech or any indication of anomalous variance for Student 8.

Overall, only Student 20 from the explicit group experienced a clear phase transition (change) in her speech fluency development, as indicated by the three criteria of analysis. The other three participants did not show the same indications. These results reveal the difference in L2 speech fluency development on an individual basis. The development could also be different between students who received the same instruction and practice opportunities, as shown by the varied results for each of the two participants from the same groups in the study.

Since not all participants' data showed conclusive results of phase transition, the phase transition experienced by one (Student 20) might not have been caused by the instruction. The phase transition could only result from participants' individual development in their oral performance, especially in producing increasingly fluent language. Meanwhile, the participants' affective factors, such as motivation, beliefs, and self-efficacy, might trigger qualitative changes in the development. These factors may play an important role as agents of change within the participants' linguistic system (Baba & Nitta, 2014). Therefore, for more comprehensive results in the analysis of phase transitions, more evidence and techniques were required to prove the existence of a sudden change in language development.

Chapter 9: Discussion of the Results

In this chapter the results of two analyses performed in two previous chapters are discussed. Section 9.1 highlights the overall findings for the study, following each research question. The implications of these findings for theories and practice in L2 instruction are discussed in Section 9.2. Section 9.3 is an outline of the limitations of the study and directions for future research. Finally, Section 9.4 provides the conclusion of the study.

9.1 Findings of the study

The first research question asked whether explicit fluency strategy training instruction would influence L2 learners' oral fluency. Despite the lack of the statistical significances, the results revealed that the explicit strategy training instruction may have had a potentially meaningful impact on certain aspects of participants' speech fluency (see Section 7.3.1). To be specific, the explicit strategy training could potentially improve participants' phonation time ratio and mean duration of pauses (breakdown fluency), with medium effect sizes evident. At the same time, the improvement in speech rate (speech fluency) was small. Improvements in these measures indicate that the participants had a developmental shift from speaking shortly and slowly with long and frequent pausing to a slightly longer and faster speech with shorter and fewer pauses.

To some extent, these findings are aligned with the results of similar previous studies investigating the effects of fluency strategy training on L2 fluency, such as Tavakoli et al. (2016) and Seifoori and Vahidi (2012). These researchers found that fluency strategies training significantly improved their participants' speech fluency, based on some fluency measures. Information regarding the effect sizes can only be found in Tavakoli et al.'s (2016) study. To be specific, Tavakoli et al. (2016) found that their participants improved on phonation time ratio, the mean number of pauses clause-internal, the mean length of run, speech rates, and articulation rate. Except for speech rate with a medium effect size, other variables that significantly differed between the pre-test and post-test scores only had small effect sizes.

The current study found more substantial effects than Tavakoli et al.'s (2016) study on phonation time ratio and mean duration of pauses. On the contrary, Tavakoli et al. observed a more substantial effect on articulation rate. The effects of other variables are comparable, including repair measures that failed to show improvement. Improvement in the speech rate, articulation rate, and mean length of run experienced by Tavakoli et al.'s participants indicates that they could produce more syllables with fewer pauses in speech as a result of the short (four week) strategy training intervention. Meanwhile, the speech fluency improvement in terms of phonation time ratio and mean duration of pause, as shown by participants in the current study, suggests that the eight-week strategy training intervention enabled the participants to produce speech with more syllables and the same number but shorter pauses.

One distinct difference between the current study and the two previous studies was the participants' English proficiency level. Tavakoli et al.'s study involved English for Academic Purposes learners from a university in the United Kingdom. The learners were identified as B2 level according to the Common European Framework of Reference (Little, 2006) with IELTS 5.0-5.5 or equivalent. Meanwhile, Seifoori and Vahidi (2012) selected their participants from intermediate level English learners in Iran. In Tavakoli et al.'s study, the higher proficiency of the participants with extensive exposure to the target language inside and outside the university might have increased the strategy training's effect. Similarly, Seifoori and Vahidi's (2011) results might have been influenced by the participants' proficiency level. Although their participants did not live in the target language environment, they were high-scoring learners among the intermediate English learners at the university.

In comparison, the current study involved participants from a third-tier university in Indonesia whose proficiency level ranged from A2 to B1 (see Section 6.3 for details of the participants). EFL learners' intact classes were used, while in the two previous studies, groups of intentionally selected learners were involved. While these researchers could selectively choose their participants based on the English proficiency level, the current study had to accommodate students with different levels due to the classroom formation.

However, despite the differences in proficiency level and learning environment, the participants in the current study could still make a more meaningful improvement than those in Tavakoli et al.'s (2016) study in terms of the effect sizes in some of the speech fluency measures. To some extent, this finding could indicate the advantage of estimation analysis over the NHST.

For EFL learners in the current study, a minor improvement in their speaking skills could be considered meaningful. These learners do not have exposure to the target language environment to practise their learned skills outside the classroom. Speaking needs actual and repeated practice to be perfected. However, these learners do not have real practice opportunities because English is not used in their community. Therefore, if the learners can improve their speech slightly, this could be essential to their language skills development. A minor improvement could be recognized easily by their teachers or peers. A shift from speaking slowly with much pausing to slightly faster with fewer and shorter pauses should be noticeable. Therefore, it can be argued that, within the context of the present study, the strategy instruction had an essential impact on EFL learners' speech fluency.

The second research question asked whether implicit task-based instruction would influence L2 learners' speech fluency. The results indicated a noticeable improvement in the articulation rate which is one aspect of the speed fluency (see Section 7.3.2). The articulation rate improved with a small effect size. Meanwhile, improvements on other measures, such as the speech rate and mean duration of pauses, were negligible. This result indicates that the implicit instruction with a repeated practice opportunity could slightly improve learners' speech in terms of speed. For participants in the current study, improvement with a small effect size could be considered essential because of their limitation as third-tier university students in an EFL context.

This finding was different from several previous studies that revealed the beneficial effects of task repetition (e.g., Ahmadian & Tavakoli, 2011; Bygate, 2001a; Gass, Mackey, Alvarez-Torres, & Fernández-García, 1999; Lambert et al., 2017; Wang, 2014). Among these studies, only Lambert et al. (2017) and Wang (2014) reported the effect sizes of the results.

Specifically, a large effect size on speech rate was reported as the impact of task repetition. Meanwhile, in the current study only a negligible effect size on speech rate was evident. This difference could be due to different methodological aspects and participants' different levels of English proficiency. For example, Lambert et al. (2017) allowed their learners to repeat the same tasks six times with different interlocutor, finding that a significant improvement in fluency occurred until the fourth repetition of the task. Participants in this study repeated the same task six times as the speaker and six times as the listener. Therefore, they could also benefit from their interlocutors' performances.

Meanwhile, Wang's (2014) participants were learners with medium to high English proficiency with IELTS scores ranging from 6 - 7.5. Wang's participants repeated the task twice: they watched a video and narrated the story simultaneously two consecutive times. In addition, both studies were conducted in a 'laboratory' environment with a designed setting and selected participants, which was different from an actual English classroom environment. Therefore, the results from both studies could be subject to some doubt in terms of ecological validity since language instruction should include real language learners in language classrooms (Doughty, 2003).

In addition, the current study also attempted to compare the impact of the two instructional conditions on L2 speech fluency. The results showed no statistical significance and meaningful difference between the impacts of both instructional conditions in five speech fluency measures. This implied that implicit task-based instruction was comparable to explicit strategy training instruction in developing EFL learners' speech fluency in the current study's context. Learners who learned and practised strategies to improve their speech fluency exhibited, to some extent, similar development to those who only received task practice (FonM). Each instruction also seemed to impact different aspects of learners' speech fluency. The explicit strategy instruction improved learners' breakdown fluency (phonation time ratio and mean duration of pauses) more than the implicit strategy training did. The impact of implicit instruction was more substantial than explicit instruction on the speed fluency (articulation rate). Meanwhile, the effect of both instructional conditions on the speech rate was highly comparable. The finding regarding speech rate improvement

corroborates with those of previous studies (e.g., Kormos & Denes, 2004; Lennon, 1990; Towell et al., 1996) which conclude that the speech rate is strongly associated with the development of L2 speech fluency.

To date, no researchers have attempted to compare explicit fluency strategy training instruction with implicit instruction. Therefore, the results of the current study cannot be directly compared with any other study, however, Sato's (2020) study seems to be similar to the current study for comparative purposes. While the current study focused on investigating the impact of explicit strategy training instruction, Sato (2020) investigated the impact of explicit metacognitive instruction for collaborative interaction (MICI) during communicative tasks. Sato compared metacognitive instruction with implicit task-based instruction. Explicit metacognitive instruction was given to one group of learners, while the other group performed the assigned tasks through implicit instruction. In Sato's study, both instructional groups were supplied with communicative tasks, and learners' achievement was measured based on their comprehensibility and the use of communicative strategies. Sato found that learners who were given metacognitive strategy-based instruction outperformed those with implicit task-based instruction in strategy use and comprehensibility. As measured by effect size (Cohen *d*), the magnitude of differences between the two forms of instruction was large. Since Sato found that explicit strategy training via metacognitive instruction might be practical to improve learners' communicative strategy use, it would be interesting to verify whether the instruction could also improve learners' fluency strategy use in a similar fashion.

The third research question concerned the development of L2 learner's oral fluency under the two instructional conditions over a ten-week duration. The ten time series data of four high-achieving participants, two from each group, were analyzed to determine phase transitions. The results showed that each participant had a different developmental pattern. One participant from the explicit instruction group showed a clear phase transition in her speech fluency development, as represented by the number of syllables in each speech performance. This finding was supported by multiple methods of phase transition analysis: the sudden jump, anomalous variance, and a qualitative change in the speech. These

multiple methods of analysis are required to convincingly claim the existence of a phase transition (Baba & Nitta, 2014). The other participant in the explicit group performed differently, without any clear indications of a phase transition.

Meanwhile, neither of the two participants in the implicit instruction group experienced a substantial change or phase transition in their speech fluency development. Although the change point analysis results on the participants' time series data indicated a phase transition at the end of the ten weeks of instruction, no supporting evidence was found in the anomalous variance and qualitative change analysis. The results of the phase transition analysis for the four participants are summarised in Table 9.1.

Table 9.1 Summary of phase transition analysis results of the four participants

Analysis	Explicit Group		Implicit Group	
	Student 20	**Student 18**	**Student 8**	**Student 22**
Sudden Jump	Yes	No	Yes	Yes
Anomalous Variance	Yes	No	No	No
Qualitative Change in the Attractor	Yes	No	No	Partially
Overall Phase Transition	**YES**	**NO**	**NO**	**NO**

Note. A confirmed phase transition must be supported by all three analyses.

The four participants involved in the analysis had high speech fluency gains in the four speech fluency measures. However, they were different in terms of background and proficiency level as measured by the TOEIC test (see Section 8.1). Student 20, who showed a clear phase transition, was categorized as a high proficiency learner, with a TOEIC score of 635. Since the other participant with high proficiency (Student 8, with a TOEIC score of 655) did not experience a phase transition, the result shown by Student 20 might not have been a direct effect of her proficiency level.

In addition, both of these highly proficient participants had different anxiety levels, as revealed by the motivation questionnaire administered at the beginning of the study, where Student 20 had higher anxiety than Student 8. Here Student 20 indicated that she would be nervous and anxious when dealing with the English lesson, especially when speaking in front of others. In contrast, Student 8 indicated she had more confidence and a comfortable

feeling about English. Therefore, learners' anxiety level could be considered as a non-influential factor for the instruction within the context of the current study.

In terms of the variability within and between participants, this study's results are similar to those of Larsen-Freeman's (2006) and Baba and Nitta's (2014) studies, in which individual participants' production did not follow the same developmental pathway. However, considering the context similarity, this study's result differed from a similar phase transition analysis study conducted by Baba and Nitta (2014). Baba and Nitta (2014) found that the two participants indicated strong phase transitions in developing their writing fluency as measured by the number of words produced. They chose their two participants from an intact class, based on their weekly writing tasks. These students had the highest correlations between the week number (30 weeks in total) and their weekly writing fluency, which also meant the greatest change in fluency. Baba and Nitta employed four methods of analysis: sudden jumps, anomalous variance, divergence, and qualitative change in the attractor. Therefore, the differences between their study and the current study were the method of choosing the participants, the method of analyzing the phase transition, and the number of instructional meetings.

Unlike Baba and Nita's study, which only involved one variable (number of words), in the current study, six variables were used to measure learners' speech fluency. Therefore, instead of correlational analysis, the present study chose the participants based on their speech fluency gains on the four measures, selecting those with the most considerable improvement (see Section 8.1 for details). Despite this difference, both selection methods served the same function: to choose participants with the greatest fluency changes. Meanwhile, 'divergence' was the only difference in the analysis of the phase transition when comparing Baba and Nitta's and the current study. The current study could not apply divergence because there was no interview data or reflective comments from the participants for analysis purposes. The absence of divergence analysis could be one limitation of the present study. However, the availability of the other three methods of analysis can be considered adequate to detect and confirm a phase transition.

The number of instructional meetings was predicted to be the main cause of the different results between the current study and Baba and Nitta's study. Baba and Nitta found a clear phase transition in their participants' writing fluency development after a long period of instruction (30 weeks, i.e., one academic year). One participant's change point occurred in Week 18, while the other experienced this in Week 25. The current study was conducted for ten weeks due to limitations of time and available funds. Within this duration, three participants indicated a phase transition in their speech fluency development, between Week 7 and Week 10. Maybe, if the study were conducted for longer, more phase transitions with a more precise indication could also be found.

This phase transition analysis was intended to explore learners' speech fluency development on an individual basis. Therefore, it would not be advisable to relate the results to the impact of pedagogical instruction. The fact that only one participant showed a clear phase transition in her speech fluency development could be due to different variables, or at least a combination of different variables, instead of the instruction alone. First, the phase transition could be caused by the participant's individual development in producing fluent language. Second, the qualitative changes in her speech production might result from her affective factors, such as motivation, beliefs, and self-efficacy, as agents of change within her linguistic system. It is impossible to make any claims regarding the impact of instruction based on the phase transition analyses only, especially when not all participants showed conclusive results.

The fourth research question was whether the two instructional conditions affect learners' oral proficiency as measured by teachers' ratings. The results showed that learners in both groups had meaningful improvement in their oral proficiency. The effects of both instructional conditions were also found to be comparable. Each instruction form seemed to have stronger effects on different oral proficiency aspects. The explicit strategy instruction had a better impact on learners' general proficiency, vocabulary, and complexity. Meanwhile, implicit instruction seemed to have a better impact than explicit instruction on learners' pronunciation, discourse, and grammar. Therefore, it can also be argued that implicit instruction could improve learners' oral performance, albeit slightly differently.

Although initially intended for influencing L2 speech fluency, based on the teacher's perception, both instructional conditions improved learners' general oral proficiency. These findings corroborated previous research, such as the meta-analyses conducted by Norris and Ortega (2000) and Spada and Tomita (2010). These two meta-analyses found that classroom instructions significantly affect L2 oral performance. However, most studies reported in those meta-analyses measured the effects of instruction based on learners' improvement in using specific language features, such as simple and complex forms. No researcher has tried to measure the effects of instruction by using raters' judgements. The current study's findings suggest that the use of raters' judgements might also be necessary for classroom-based studies to measure the effect of instruction on gains in learners' oral proficiency.

To date, raters' judgement is only applied to score learners' oral performance based on specific criteria or scales. Studies on L2 oral performances with rater's judgements usually focus on investigating the variability among raters (e.g., Bachman, Lynch, & Mason, 1995; Isaacs & Thomson, 2013; Winke, Gass, & Myford, 2013), including the influence of raters' accent on ratings (e.g., Carey, Mannell, & Dunn, 2011; Derwing & Munro, 1997; Huang, Alegre, & Eisenberg, 2016). Therefore, the findings of the current study offered an alternative function of rater's judgment for classroom-based instruction.

9.2 Pedagogical implications of the study

9.2.1 The comparable and complementary impact of both instructional conditions on learners' speech fluency

For EFL classrooms in the context of the current study, explicit strategy training and implicit task-based instruction are potentially comparable as well as complementary in terms of affecting learners' speech fluency. Each instruction type could impact different aspects of learners' speech fluency. The explicit strategy training works better than the implicit instruction on improving learners' breakdown fluency, while the implicit instruction was better in speed fluency. It seems that the impacts of both instructional conditions are complementary for EFL classroom instruction within the context of the study. Therefore, the

combination of both instructional conditions could have a stronger impact on improving EFL learners' speech fluency.

The improvement in breakdown fluency experienced by learners under the explicit strategy training could be connected to fluency strategies regarding pausing and the task repetition practice (see Section 6.5.1.2). During the first three meetings of the eight-week strategy training, learners were trained to handle their pausing during speech. It seems that, after the instruction, this training on pausing had a meaningful impact on learners, so they applied it in their speech in the form of fewer and much shorter pauses. As a result, the phonation time ratio and mean duration of pauses in the speech improved in the post-test.

In the implicit instruction, learners' improvement in speed fluency could be linked to the repeated practice (task repetition) condition. Learners under this implicit instruction were given more practice time than the explicit group at the task preparation stage (the second stage) of the instruction (see Section 6.5.1.2.B). Learners in this implicit group practised 'a similar task' six times in addition to the practice of 'the same task' three times. This additional similar task repetition might have enabled learners in the implicit instruction to become more familiar with the task they performed in the following stage than learners in the explicit instruction. This familiarity may have enabled learners to understand "basic procedures associated with the task type, and improve their ability to adapt L2 resources for task-relevant functions" (Lambert, Aubrey, & Leeming, 2020, pp. 25-26). This task familiarity seemed to help learners to speak slightly faster after the instruction, as reflected by a minor improvement in the articulation and speech rates in the post-test.

The pedagogical implication of the current study regarding the complementary impact of both instructional conditions supports Long's (2017) claim concerning EFL classroom instruction and learning tasks. Long (2017) argues that EFL classroom instruction should not focus only on a single instruction to gain the full benefits of the language learning task. "Neither explicit nor implicit learning, alone, will suffice, especially in foreign language settings, so the solution will have to be something else" (Long, 2017, p. 20). Therefore,

whenever learning tasks are utilized, a combination of explicit and implicit instruction might be a solution to improve the impact of classroom instruction in the EFL context.

9.2.2 The impact of both instructional conditions on the speech production process

In relation to L2 speech production, the explicit strategy training could have an impact in easing the formulation stage, which is often conscious, and helping the speech production stages to work in parallel. This conscious and unparallel mechanism may result in a slow formulation process and a breakdown in speech production (Kormos, 2006; Levelt, 1989), which is characterized by grammatical and lexical errors, repetition of the same words with or without modification, and pauses, either filled or unfilled, within and between clauses (de Jong & Perfetti, 2011).

Since learners were able to improve their breakdown fluency, it is argued that they have proceduralized their phonological and grammatical rules to become procedural knowledge (see Section 2.1.1). The procedural knowledge enables learners to automatically perform the encoding process quickly and in parallel with other processes because procedural knowledge puts less pressure on the capacity-limited working memory (de Jong & Perfetti, 2011). Hence, their speech would be more fluent because learners do not need to pause more or longer to search for the required lemmas in their mental lexicon, including the syntactic, phonological, and morphological information for the lemmas. All the information has been stored in the form of procedural knowledge and is ready to use to produce speech.

Meanwhile, the implicit instruction seemed to help learners produce a slightly better L2 speech after the instruction. The improvement in the speed of speech measures under the implicit instruction suggests that their speech production mechanism, from conceptualization to articulation (see Section 1.1), becomes more efficient (Kormos, 2006). It seems that the benefits offered by the task being practised six times (in the form of task familiarity) increased the impact of the following three-time task practice.

Skehan (2014) argues that a macro-structure of the conceptual plan for the message to be conveyed is built at the first performance of the three-time task practice. The first performance also eases the linguistic formulation of a message by providing syntactic,

morphological, and phonological information, and activating syntactic building procedures. The articulation stage is also helped through the first performance since the phonetic samples to be used to transform the phonetic plan into overt speech are available at this stage. Once all these three major stages are processed in the first performance of a task, the second and the third performance could be done faster. Therefore, in the context of EFL learners in the current study, the combination of both types of task repetition seemed to enable learners to produce a speech efficiently, which resulted in improved speech fluency.

9.2.3 The impact of both instructional conditions on L2 oral proficiency

Both instructional conditions in the current study were found to have a meaningful impact on improving learners' oral proficiency. Each instruction had a stronger impact on different aspects of oral proficiency. The explicit instruction seemed to highly affect learners' general proficiency, vocabulary, and complexity. The implicit instruction also highly influenced learners' oral proficiency. Meaningful impacts were specifically found on pronunciation, discourse, and grammar.

It seems that instructing participants to explicitly focus on aspects that could improve their speech fluency, such as avoiding and filling their pauses, and avoiding making repetitions and repair moves, could improve oral performance, resulting in a better organized speech with varied vocabulary and some embedded clauses. Mostly, learners' skills under the explicit instruction improved by one level from Level 2 (beginning speaking) to Level 3 (competent speaking).

When instructed implicitly, learners were able to produce a more plausible and logically informed speech. The speech also had fewer phonemic and grammatical errors and better intonation patterns. The skill improvement of learners in this group was still within the same level (Level 3). However, the range of improvement (mean difference between the pre-test and post-test) is almost similar to that of the explicit instruction. This also indicates that the implicit group had a slightly higher starting point than the explicit group in oral proficiency.

These results suggest that the use of raters' judgements might also be necessary for EFL classroom-based studies to measure the effect of instruction on gains in learners' oral

proficiency. EFL classrooms are known to be heterogeneous in terms of language proficiency; therefore, generalizing the results of an instructional activity is challenging. Standardized tests or measurements used in EFL classroom research might only be applicable for learners with specific proficiency levels but not for all learners in the same class. EFL teachers could deal with this heterogeneity because they know their learners, and they could be sensitive to changes or improvements in learners' proficiency.

EFL teachers usually know and understand their learners' language skills, especially from daily classroom interaction. These teachers are also sensitive to students' developmental change as the result of the learning. They are familiar with end-of-course assessments or other classroom final semester tests conducted regularly. Therefore, teachers with this experience could be involved in assessing learners' oral proficiency development due to instruction related to perceived fluency rather than cognitive and utterance fluency (Segalowitz, 2010).

Furthermore, teachers' subjectivity in assessing or rating learners' performances could be reduced by providing reliable rating scales and training them on how to apply the scales. Teachers could also be informed and assured about the specific purpose of the rating to assess learners' development objectively. In addition, the teaching experience possessed by these teachers offers another advantage in the rating because raters with teaching experience may be more focused on language pragmatics, content, and rhetorical organization than the surface language features (Cumming, 1990; Kim, 2009). These factors would ensure the accuracy and objectivity of the rating results.

Creating an oral performance rating scale could be costly and time-consuming. However, reliable rating scales are readily available. A combination of holistic and analytic ratings should be used (Bachman & Savignon, 1986) to provide a detailed oral proficiency profile for each learner. A holistic rating captures the overall impression of oral performance. In contrast, an analytic rating assesses various performance categories, such as content, delivery, organization, and language features. An analytic rubric is used to identify language subskills, such as grammar, vocabulary, pronunciation, and fluency (Fulcher, 2003). For this

purpose, the combination of a holistic rubric provided by Brown (2001) and an analytic rubric employed by Choi (2005) could be implemented since these ratings are widely used and proven to be reliable in assessing L2 oral proficiency.

Overall, in the present study, it was found that teachers' judgements could be essential in classroom-based studies investigating L2 oral proficiency. For this purpose, the combination of a reliable holistic and analytic rubric to rate learners' improvement would be required to obtain valid and detailed information.

9.2.4 Variability in learners' speech fluency development

As found in many studies based on the complex dynamic systems theory, in the current study some variability in the development of speech fluency was also found within one individual learner and among learners. In complex dynamic systems theory, variability is regarded "as a potential driving force of development and a potential indicator of ongoing processes" (Van Geert & Van Dijk, 2002, p. 341). The variability within an individual learner can provide insight into the developmental dynamics of L2 learners (Verspoor et al., 2008). Furthermore, as Verspoor et al. (2008) explained, that this within individual variability occurs continuously, and the degree of variability may change depending on the system's stability at a particular moment. If the system enters a relatively more unstable period, the variability will increase. Increasing variability suggests that the system is changing. "By looking at the different degrees and patterns of variability in dense developmental data, we can discover how and when different subsystems are changing and developing, and how they relate to each other" (Verspoor et al., 2008, p. 215).

The finding of this study regarding the variability in learners' development can strengthen the pedagogical point of view regarding learners' learning development at the individual level. Individual learner in the study showed different development or improvement. In fact, one of the participants experienced a developmental shift in her speech fluency, entering a mode of higher development. This phase shift occurred after she showed high variability in her developmental data. Hence, this finding is consistent with the findings of previous studies under complex dynamic systems theory (e.g., Baba & Nitta, 2014; Fogal, 2020;

Larsen-Freeman, 2006; Lowie & Verspoor, 2019; Verspoor et al., 2008), where within and among individual variability in second language development were also found.

9.2.5 Complementary methods of analysis

Based on the overall findings of the current study, it could be concluded that in investigating EFL oral performance, group studies with inferential and estimation analysis and individual case studies under complex dynamic systems theory are complementary. While group studies generate "valuable information about the relative weight of individual factors that may play a role in L2 development, longitudinal case studies are needed to understand the process of individual learners' development" (Lowie & Verspoor, 2019, p. 184). Investigating individual learner's development and variability surrounding the development could be helpful to provide complementary information about the process of development. When a group study failed to achieve a conclusion due to the lack of significance results, the case study could still explain the actual situation on individual learner's development. The overall findings of the current study suggest that even though an instruction did not significantly impact learners in a group level, the impact could still be influential in individual learners. Hence, the information regarding the influential impact of instruction on individual learners could be used to modify the instruction so the impact on group level could be improved.

9.3 Limitations of the study and recommendations for future research

The current study had some limitations that could not be overcome during the research period. These limitations shape some of the recommendations for future research. The first limitation was the lack of an independent control group that completed the pre-test and post-test without either implicit or explicit instruction. Data from a control group would provide a baseline to compare with the two instruction groups' data. As the present research was conducted as a classroom-based study in a university language program, a control group was not possible for ethical reasons. Learners in the program could not be deprived of instruction. Hence, it is recommended that future research include an independent control group to obtain more comprehensive results.

The second limitation of the study was the absence of post-task activities. It is recommended for future research to include post-task activities in the instruction because providing learners with these could improve the effects of explicit strategy training. Skehan (2007) argues that post-task activities provide learners with subsequent consequences, which could influence them to apply a different approach to task performance and somewhat focus on what was instructed (focus on form). One post-task activity that could be applicable for the context of the study is assigning the students to analyze the recording of their own task performance to find proof of the various strategies they used, such as fillers, pausing, or other breakdown and repair fluency indicators like repetitions, hesitations, and false starts. Hence, it is hoped that the students will recognize their own patterns and habits of performing a speech and make appropriate improvements for their next speech performance.

The third limitation of this study was due to the time constraint in conducting the experiment which could only take place for a total of ten weeks, including the pre-test, the instructional meeting, and the post-test. Strategy training requires more practice, which, in EFL classrooms, constitutes more meetings or weeks. The instruction could also be applied with short classroom meetings for highly proficient learners in an ESL, not EFL, environment (e.g., Tavakoli et al., 2016), or with selected highly proficient EFL learners (e.g., Seifoori & Vahidi, 2012).

In addition, the phase transition analysis under the complex dynamic system approach might also be more effective when a longer data series is used. A data series constitutes the number of meetings, normally one a week during a semester or an academic year in the EFL classroom. As revealed in this study, three learners showed signs of a phase transition at the end of the ten week data series. Therefore, an analysis comparing the data before and after the phase transition would not be possible. A longer series might have given them time to develop a more valid phase transition and hence this is recommended for future research.

In future classroom research, it is also suggested to have more instructional meetings than occurred in the current study which had only one intervention and eight weekly meeting for

experimentation. More meetings will give learners more time to practise the learning strategy and sufficient time as well as data to develop phase transitions in their learning development. At least one semester long study consisting of 14 to 16 meetings should be prepared for future research. More meetings would enable more meaningful results to be generated from the analysis.

The fourth limitation deals with the post-treatment questionnaire. For a study involving strategy training, a post-treatment questionnaire would be useful to reveal the participants' use of the strategy training. This information helps show the actual condition of the student's understanding of the strategies, which could be compared with the study results. This information could only be collected at the end of the study in the form of a questionnaire or an interview. Usually, a questionnaire's questions are directed at the participants' practice in using the strategies they have learned during the instruction. Unfortunately, this study did not apply the post-treatment questionnaire because the main objective was to collect data regarding the effects of the two forms of instruction, not the strategy use. Therefore, it is suggested for future research to administer a post-treatment questionnaire.

Additionally, EFL teachers could use a combination of explicit and implicit instruction to improve learners' speech fluency comprehensively. The individual application of each form of instruction could potentially improve specific aspects of speech fluency. Therefore, a combination could potentially offer a more rounded improvement.

9.3.1 Improving automatization of the speech production process

A recommendation is also proposed to improve the effectiveness of both instructional conditions on the automatization of the speech production process. It is recommended that future researchers use the same task for the two oral production stages (the task preparation and output-based task stages) in the instruction. To avoid boredom because of doing the same task repeatedly (Bygate, 2001a), the present study involved different tasks for the second (task preparation) and third (output-based) stages of the lesson (see Section 6.5.1). It was hoped that learners would still gain some sense of task familiarity, which could

benefit later performance. However, using different tasks seemed to give learners another problem because they had to watch a new video, make new notes, and conceptualize the content of the speech for the performance. With this level of learners, it might be beneficial to use the same task repetition instead of parallel tasks (Lambert et al. 2020). Previous research has shown that even at the intermediate level, EFL learners who have little chance to practise the language outside of the classroom do not report being bored when repeating the same task up to six times (Lambert et al., 2017). In fact, they can appreciate the practice opportunities of such task repetition.

In addition, it is also hoped that using the same task repetition for the task preparation and output-based stages in the instruction would enable learners to fully proceduralize their linguistics knowledge. Proceduralization usually occurs during the encoding stages of language production (Kormos, 2014) and could be achieved through many encounters with the same linguistic items (de Jong & Perfetti, 2011) (see Section 2.1.1). The use of the same task would allow learners to incorporate certain utterances and structures from a previous practice into the next one, so they would not need to spend more time (pausing) to encode the intended message.

9.3.2 Improving the fluency strategy training

Another recommendation is to improve the fluency strategy training. To enhance strategy use, learners need more practice opportunities to apply these strategies and become more familiar with such strategy use. In this study, a different strategy was learned and practised by learners in every meeting. Therefore, learners did not have the opportunity to practise and apply the learned strategy in the next meeting or one week after their first encounter with the strategy. Having the opportunity to practise the strategy in the next meeting could increase learners' opportunity to enhance their speech fluency. Hence, when learners can constantly apply the strategy to their oral performance, they may produce more fluent speech with a decreasing number and duration of pauses. They would know how to avoid silence during the speech or fill their potential silence with lexical fillers.

In addition, the explicit strategy training could also be improved by introducing a reactive focus on form (Long, 1991) instruction. The strategy training applied in the present study provides learners with a proactive focus on form in which learners' attention was directed to aspects of utterance fluency before practising the task. However, there was the possibility that learners may have forgotten the instruction during the task practice. Therefore, their performances did not specifically follow the instruction given to them. A reactive focus on form could help learners concentrate on the instruction because teachers could directly react to their practice whenever needed by reminding them to use the strategy. For example, when learners seemed to pause longer during the speech, teachers could remind them to use the lexical fillers instead of just being silent. This reactive focus on the form might further raise learners' awareness of the strategy use; therefore, they might be encouraged to increase strategy use during the task performance.

9.4 Conclusion

The primary objective of this study was to investigate the effects of two instructional conditions in the classroom (explicit fluency strategy training and implicit task-based instruction) on the development of the speech fluency of low to intermediate level Indonesian learners of English as a foreign language at a third-tier university in Indonesia. The secondary aims of the study were to explore the development of speech fluency of individual learners and investigate the effects of the instruction on learners' oral proficiency, based on raters' judgement.

Two intact classes of year English were involved in the study. Each class received different instruction but performed the same news presentation tasks orally. Quantitative comparison and complex dynamic system approaches were adopted to collect and analyze data, employing two instruments – the recorded oral performance and the self-report questionnaire.

The inferential and estimation analyses showed that both instructional conditions demonstrated comparable trends towards gains in EFL learners' speech fluency. In terms of utterance fluency construct, the explicit strategy training could potentially have a more

meaningful impact than the implicit instruction on learners' breakdown fluency; the implicit instruction's impact seemed to be better in speed fluency. These findings also indicate a complementary impact of both instructional conditions for EFL classroom instruction. A combination of both instructional conditions could potentially be applied in EFL classroom to have a better impact on improving EFL learners' speech fluency. In perceived fluency, both instructional conditions also impacted learners' oral proficiency significantly and meaningfully, as revealed by teachers' ratings. These findings point to a possible difference in the underlying constructs of perceived fluency, on the one hand, and cognitive/utterance fluency, on the other.

As the secondary aim of the study, a phase transition analysis using the complex dynamic system approach was conducted. This revealed that each learner had a different developmental pattern, confirming EFL learners' inter- and intra-individual development. One learner from the explicit strategy instruction group experienced a sudden jump in her development of speech fluency, indicating that she entered a new mode of speech fluency development during the study. Meanwhile, the other three did not show a transition to a higher developmental phase. These results may have been due to the limited time scale of the study, as these other three learners demonstrated initial moves towards a phase transition towards the end of the study. Nevertheless, since only one participant showed a clear indication of one phase transition, as discussed in Section 9.1 above, a combination of different variables instead of just the instruction could be the cause of this single successful phase transition.

In addition, the study also revealed that learners' individual differences in terms of proficiency level, motivation, and self-regulation had no relationship with the development of speech fluency. Therefore, for EFL learners in the context of the present study, the impact of both instructional conditions was not affected by their academic and socio-affective backgrounds.

Based on the overall results of the study, it could be concluded that in investigating EFL oral performance, group studies with inferential and estimation analysis and individual case

studies under complex dynamic systems theory are complementary. The results also suggest that both instructional conditions might be more productively implemented in EFL classrooms to improve learners' speech fluency and oral proficiency, depending on the context, provided certain modifications are made in how these conditions are implemented. Those modifications are (1) using the same as opposed to parallel task repetition, (2) providing more opportunities to practise the learned strategy, (3) applying a reactive focus on form instruction for the strategy training practice, (4) implementing post-task activities, and (5) implementing semester long (14-16 weeks) instructional meetings or data series.

With these modifications, both instructional conditions might be applied as an alternative to traditional structure-based PPP approaches in EFL classrooms in which learners' oral performance is a priority. Finally, a combination of both instructional conditions could potentially offer a better impact on EFL classroom learning.

References

Abraham, R. (1994). *Chaos, Gaia, Eros a Chaos Pioneer Uncovers the Three Great Streams of History*. New York: Harper-Collins.

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research, 15*(1), 35-59.

Anderson, J. R. (1983). Cognitive science series. The architecture of cognition. Hillsdale, NJ, US. In: Lawrence Erlbaum Associates, Inc.

Anderson, J. R. (2000). *Learning and memory: An integrated approach*: John Wiley & Sons Inc.

Arczis Web Technologies, I. (Producer). (2019, 2019). Syllable Count. Retrieved from http://www.syllablecount.com/#results

Ardasheva, Y., Wang, Z., Adesope, O. O., & Valentine, J. C. (2017). Exploring Effectiveness and Moderators of Language Learning Strategy Instruction on Second Language and Self-Regulated Learning Outcomes. *Review of Educational Research*, 0034654316689135.

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning, 64*(1), 1-35.

Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing, 12*(2), 238-257.

Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *The Modern Language Journal, 70*(4), 380-390.

Batty, A. O., & Stewart, J. (2016). Theoretical Underpinnings of a Computerized Objective Test of Communicative Competence in Japan. In V. Aryadoust & J. Fox (Eds.), *Trends in Language Assessment Research and Practice: The View from the Middle East and the Pacific Rim* (pp. 388). Newcastle, UK: Cambridge School.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., . . . Larsen-Freeman, D. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*, 1-26.

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and instruction, 7*(2), 161-186.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.09. Retrieved from http://www.fon.hum.uva.nl/praat/

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.

Brown, H. D. (2006). *Principles of language learning and teaching* (5th ed.). New York: Pearson Education Ltd.

Brown, J. D. (2001). Pragmatics tests: Different purposes, different tests. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 301-325). New York: Cambridge University Press.

Bui, G., Ahmadian, M. J., & Hunter, A.-M. (2019). Spacing effects on repeated L2 task performance. *System, 81*, 1-13.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic research, 4*(1), 75-87.

Bygate, M. (1996). *Effects of task repetition on language structure and control.* Paper presented at the AAAL Conference, Chicago.

Bygate, M. (2001a). Effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic task: Second language learning, teaching, and testing.* (pp. 23-48). Harlow: Longman.

Bygate, M. (2001b). Speaking. In D. Nunan & R. Carter (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 14-20). Cambridge: Cambridge University Press.

Byrne, D. (2002). *Interpreting quantitative data*. London: Sage.

Byrne, D. (2005). Complexity, configurations and cases. *Theory, culture & society, 22*(5), 95-111.

Byrne, D., & Ragin, C. C. (2009). *The Sage handbook of case-based methods*: Sage Publications.

Cameron, L. (2015). Embracing connectedness and change: A complex dynamic systems perspective for applied linguistic research. *AILA Review, 28*(1), 28-48.

Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing, 28*(2), 201-219.

Chafe, W. (1980). Some reasons for hesitating. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 169-180). De Haque Mouton.

Chamot, A. U. (2004). Issues in language learning strategy research and teaching. *Electronic journal of foreign language teaching, 1*(1), 14-26.

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual review of applied linguistics, 25*, 112-130.

Chamot, A. U., Barnhardt, S., El-Dinary, P. B., & Robbins, J. (1999). *Learning strategies handbook*. White Plains, NY: Addison-Wesley Longman.

Choi, I.-C. (2005a). Measurability of oral fluency through ASR-based COPI. *Multimedia-Assisted Language Learning, 8*(2), 240-261.

Choi, I. C. (2005b). Measurability of oral fluency through ASR-based COPI. *Multimedia-Assisted Language Learning, 8*(2), 240-261.

Claridge-Chang, A., & Assam, P. N. (2016). Estimation statistics should replace significance testing. *Nature methods, 13*(2), 108-109.

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England.

Cohen, A. (2003). Strategy Training for Second Language Learners. *ERIC Digest*.

Cohen, A., & Wang, I. K.-H. (2018). Fluctuation in the functions of language learner strategies. *System, 74*, 169-182.

Cohen, A., Weaver, S. J., & Li, T.-Y. (1996). *The impact of strategies-based instruction on speaking a foreign language*: Center for Advanced Research in Language Acquisition.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic press.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Academic press.

Cox, N. J. (2007). Transformation: an introduction. Retrieved from http://fmwww.bc.edu/repec/bocode/t/transint.html

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

Crookes, G. (1991). Second language speech production research: A methodologically oriented review. *Studies in Second Language Acquisition, 13*(2), 113-131.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *the Journal of the Acoustical Society of America, 111*(6), 2862-2873.

Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing, 7*(1), 31-51.
Cumming, G. (2014). The new statistics: Why and how. *Psychological science, 25*(1), 7-29.
Cumming, G. (2020). ESCI (Exploratory Software for Confidence Intervals). Retrieved from https://thenewstatistics.com/itns/esci/
Cumming, G., & Fidler, F. (2009). Confidence intervals: Better answers to better questions. *Zeitschrift für Psychologie/Journal of Psychology, 217*(1), 15-26.
Cumming, G., & Finch, S. (2005). Inference by eye: confidence intervals and how to read pictures of data. *American psychologist, 60*(2), 170.
de Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied linguistics, 13*(1), 1-24.
De Bot, K. (2003). A bilingual production model: Levelt's 'speaking'model adapted. In *The bilingualism reader* (pp. 399-420): Routledge.
de Bot, K. (2015). *A history of applied linguistics: From 1980 to the present*: Routledge.
De Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In *A dynamic approach to second language development* (pp. 5-24): John Benjamins.
de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533-568.
de Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 113-132.
de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics, 36*(2), 223-243.
de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34*(5), 893-916.
de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34*(1), 5-34. Retrieved from https://core.ac.uk/download/pdf/188465429.pdf
de Keyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195-221.
de Keyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*: Cambridge University Press.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93*(3), 283.
Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 1-16.
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*(4), 533-557.
Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54*(4), 655-679.
Devaney, R. (2003). An Introduction to Chaotic Dynamical Systems, Westview Press. *Advanced Book Program*.
Dong, J. (2016). A dynamic systems theory approach to development of listening strategy use and listening performance. *System, 63*, 149-165.
Dörnyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly, 29*(1), 55-85.

Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning, 59*, 230-248.
Dörnyei, Z. (2014). Researching complex dynamic systems:'Retrodictive qualitative modelling'in the language classroom. *Language teaching, 47*(1), 80-91.
Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language Learning, 47*(1), 173-210.
Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In CJ Doughty & MH Long (Eds.), The handbook of second language acquisition(pp. 256–310). In: Oxford: Blackwell.
Doughty, C. J. (2008). 10 Instructed SLA: Constraints, Compensation, and Enhancement. *The handbook of second language acquisition, 27*, 256.
Ellis, R. (2004). Individual differences in second language learning. In C. E. A. Davies (Ed.), *The handbook of applied linguistics* (pp. 525 - 551). Oxford: Blackwell.
Ellis, R. (2005a). *Planning and task performance in a second language* (R. Ellis Ed.). Amsterdam: John Benjamins Publishing.
Ellis, R. (2005b). Principles of instructed language learning. *System, 33*(2), 209-224.
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.
Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*: Oxford University Press Oxford.
Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London: Routledge.
Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59-84.
Europe, C. o. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*: Cambridge University Press.
Evans, D. R. (2019). *Bifurcations, Fractals, and Non-linearity in Second Language Development: A Complex Dynamic Systems Perspective.* State University of New York at Buffalo,
Field, A. (2017). *Discovering statistics using IBM SPSS statistics: North American edition*: Sage.
Fillmore, C. J. (1979). On fluency. In *Individual differences in language ability and language behavior* (pp. 85-101): Elsevier.
Finch, G. (2016). *Linguistic terms and concepts*: Macmillan International Higher Education.
Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.
Flavell, J. H., & Wellman, H. (1977). Metamemory. In R. V. Kail Jr & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Lawrence Eralbaum Associates.
Fogal, G. G. (2020). Investigating Variability in L2 Development: Extending a Complexity Theory Perspective on L2 Writing Studies and Authorial Voice. *Applied linguistics, 41*(4), 575-600.
Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299-323.
Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*(3), 215-247.
Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics, 21*(3), 354-375.
Freed, B. (2000). *Is fluency, like beauty, in the eyes (and ears) of the beholder?* Paper presented at the Perspectives on fluency.
Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing, 13*(2), 208-238.

Fulcher, G. (2003). *Testing second language speaking*. Great Britain: Pearson Education Limited.
Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*(4), 549-581.
Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22*(3), 473-492.
Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325-353.
Gilmore, R. (1993). *Catastrophe theory for scientists and engineers*: Courier Corporation.
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.
Götz, S. (2013). *Fluency in native and nonnative English speech* (Vol. 53): John Benjamins Publishing Company.
Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning, 58*(4), 747-783.
Grenfell, M. (2007). Language learner strategy research and modern foreign language teaching and learning. *Language Learning Journal, 35*(1), 9-22.
Grenfell, M., & Harris, V. (2002). *Modern languages and learning strategies: In theory and practice*: Routledge.
Gunning, P., & Oxford, R. L. (2014). Children's learning strategy use and the effects of strategy instruction on success in learning ESL in Canada. *System, 43*, 82-100.
Hassan, X., Macaro, E., Mason, D., Nye, G., Smith, P., & Vanderplank, R. (2005). Strategy training in language learning-A systematic review of available research.
Hays, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart and Winston.
Hepford, E. A. (2017). *Dynamic second language development: The interaction of complexity, accuracy, and fluency in a naturalistic learning context*: Temple University.
Huang, B., Alegre, A., & Eisenberg, A. (2016). A cross-linguistic investigation of the effect of raters' accent familiarity on speaking assessment. *Language Assessment Quarterly, 13*(1), 25-41.
Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*(2), 135-159.
Jansen, B. R., & Van der Maas, H. L. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review, 21*(4), 450-494.
Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*: MIT press.
Kim, Y.-H. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing, 26*(2), 187-217.
Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers* (Vol. 2): Association Press New York.
Kojima, M. (2001). Promoting listening strategies use in elementary English as a foreign language computer-assisted learning environment. Retrieved from https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2904&context=etd-project. Retrieved 1/09/2020, from CSUSB ScholarWorks https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2904&context=etd-project
Koponen, M., & Riggenbach, H. (2000). *Overview: Varying perspectives on fluency.* Paper presented at the Perspectives on fluency.
Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning, 49*(2), 303-342.
Kormos, J. (2006). *Speech production and second language acquisition*. New Jersey: Lawrence Erlbaum.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*, 39-59.

Kormos, J. (2014). *Speech production and second language acquisition*: Routledge.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145-164.

Kormos, J., Kiddle, T., & Csizér, K. (2011). Systems of goals, attitudes, and self-related beliefs in second-language-learning motivation. *Applied linguistics, 32*(5), 495-516.

Lambert, C., Aubrey, S., & Leeming, P. (2020). Task preparation and second language speech production. *TESOL Quarterly*.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition, 39*(1), 167-196.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied linguistics, 18*(2), 141-165.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied linguistics, 27*(4), 590-619.

Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In *Alternative approaches to second language acquisition* (pp. 60-84): Routledge.

Larsen-Freeman, D. (2018). Task repetition or task iteration? In M. Bygate (Ed.), *Learning language through Task Repetition*. Amsterdam: John Benjamins Publishing Company.

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal, 92*(2), 200-213.

Lebrun, Y. (1983). Issues in neurolinguistics. *Language Sciences, 5*(2), 241-248.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417.

Lennon, P. (2000). *The lexical element in spoken second language fluency.* Paper presented at the Perspectives on fluency.

Levelt, W. J. M. (1981). JM 1989. Speaking: From intention to articulation. *MA: The MIT Press, Cambridge*.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41-104.

Levelt, W. J. M. (1989a). Speaking: From intention to articulation. *Cambridge, MA: The MITPress*.

Levelt, W. J. M. (1989b). *Speaking: From intention to articulation. A bradford book.* Cambridge. MA: MIT Press.

Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. *Neurocognition of language*, 83-122.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38.

Lewis, M. D. (2000). The promise of dynamic systems approaches for an integrated account of human development. *Child development, 71*(1), 36-43.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language teaching, 39*(3), 167.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective, 2*(1), 39-52.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences, 20*(2), 130-141.

Lougheed, L. (2017). *Barron's TOEIC Practice Exams with MP3 CD* (3rd ed.). New York: Barron's Educational Series Inc.
Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63-88.
Lowie, W., & Verspoor, M. (2019). Individual differences and the ergodicity problem. *Language Learning, 69*, 184-206.
Mathematics, C. o. S. (2020). Dynamical System. Retrieved from https://www.math.arizona.edu/research/dynamical-systems
Meara, P. (2004). Modelling vocabulary loss. *Applied linguistics, 25*(2), 137-155.
Menn, L. (2020). Neurolinguistics. *Sociolinguistics: Linguistic Society of America.* Retrieved from https://www.linguisticsociety.org/resource/neurolinguistics
Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*(4), 610-641.
Mozgalina, A. (2015). More or less choice? The influence of choice on task motivation and task engagement. *System, 49*, 120-132.
Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied linguistics, 35*(4), 463-482.
Naiman, N., Frohlich, M., Stern, H., & Todesco, A. (1978). The good language learner: Research in education. *Toronto, Ontario, Canada: Institute for Studies in Education*.
Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal, 89*(1), 76-91.
Nation, P. (1989). Improving speaking fluency. *System, 17*(3), 377-384.
O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*: Cambridge university press.
O'Malley, J. M., Chamot, A. U., Stewner-Manzares, G., Russo, R. P., & Küpper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly, 19*(3), 557-584.
Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston, MA: Heinle and Heinle.
Oxford, R. L. (2016). *Teaching and researching language learning strategies: Self-regulation in context*. London: Routledge.
Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System, 23*(1), 1-23.
Oxford, R. L., & Crookall, D. (1989). Research on language learning strategies: Methods, findings, and instructional issues. *The Modern Language Journal, 73*(4), 404-419.
Oxford, R. L., & Leaver, B. L. (1996). A synthesis of strategy instruction for language learners. *Language learning strategies around the world: Cross-cultural perspectives*, 227-246.
Perone, S., & Simmering, V. R. (2017). Applications of dynamic systems theory to cognition and development: new frontiers. In *Advances in child development and behavior* (Vol. 52, pp. 43-80): Elsevier.
Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences, 13*(4), 707-727.
Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice, 41*(4), 219-225.
Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the meausrement of metacognition* (pp. 43.97). Lincoln, NE: Buros Institute of Mental Measurements.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: a meta-analysis. *Language Learning, 61*(4), 993-1038.
Préfontaine, Y., Kormos, J., & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing, 33*(1), 53-73.
Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning, 47*(2), 289-325.
Raupach, M. (1987). *Procedural learning in advanced learners of a foreign language*: LAUD.
Rehbein, J. (1987). On fluency in second language speech. *Psycholinguistic models of production*, 97-105.
Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition, 23*(4), 497-526.
Rigeney, J. (1978). Learning strategies: A theoretical prospective. In H. O. N. (Jr) (Ed.), (pp. 165-205).
Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes, 14*(4), 423-441.
Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes, 15*(2), 129-157.
Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*(2), 223-247.
Rubin, J. (1979). *Research strategies in the study of cognitive processes in second language learning's manuscript* University of California, Barkeley.
Rubin, J. (1981). Study of cognitive processes in second language learning. *Applied linguistics, 2*, 117.
Ruhland, R., & Van Geert, P. (1998). Jumping into syntax: Transitions in the development of closed class words. *British Journal of Developmental Psychology, 16*(1), 65-95.
Sato, M. (2020). Metacognitive instruction for collaborative interaction: The process and product of self-regulated learning in the Chilean EFL context. In C. L. R. Oliver (Ed.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 215-236). Clevedon, UK: Multilingual Matters.
Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*(4), 357-385.
Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of experimental psychology: human perception and performance, 16*(2), 227.
Segalowitz, N. (2010). *Cognitive bases of second language fluency*: Routledge.
Seifoori, Z., & Vahidi, Z. (2012). The impact of fluency strategy training on Iranian EFL learners' speech under online planning conditions. *Language Awareness, 21*(1-2), 101-112.
Seker, M. (2016). The use of self-regulation strategies by foreign language learners and its role in language achievement. *Language Teaching Research, 20*(5), 600-618.
Skehan, P. (1998). *A cognitive approach to language learning*. OXford, UK: Oxford University Press.
Skehan, P. (2003). Task-based instruction. *Language teaching, 36*(1), 1-14.
Skehan, P. (2007). Task research and language teaching: Reciprocal relationships. *Form-focused instruction and teacher education: Studies in honor of Rod Ellis*, 28-48.
Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics, 30*(4), 510-532.
Skehan, P. (2014a). Limited attentional capacity, second language performance, and task-based pedagogy. *Processing perspectives on task performance*, 211-260.

Skehan, P. (2014b). *Processing perspectives on task performance* (Vol. 5): John Benjamins Publishing Company.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93-120.

Smith, L. B., & Thelen, E. E. (1993). *A dynamic systems approach to development: Applications.* Paper presented at the This book grew out of a workshop," Dynamic Systems in Development," held for the Society for Research in Child Development in Kansas City, KS, Apr 1989.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied linguistics, 31*(4), 532-553.

Steenbeek, H., Jansen, L., & van Geert, P. (2012). Scaffolding dynamics and the emergence of problematic learning trajectories. *Learning and Individual Differences, 22*(1), 64-75.

Tavakoli, P. (2010). Pausing patterns: Differences between L2 learners and native speakers. *ELT journal, 65*(1), 71-79.

Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly, 50*(2), 447-471.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. *Planning and task performance in a second language, 239273*.

Taylor, W. A. (2000). Change-point analysis: a powerful new tool for detecting changes. In.

Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic dialogues, 15*(2), 255-283.

Thelen, E., & Smith, L. (1994). A dynamic systems approach to the development of perception and action. In: Cambridge, MA: MIT Press.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*: MIT press.

Thom, R. (1975). Structural stability and morphogenesis (DH Fowler, Trans.). *Reading, MA*.

Tonkyn, A. (2001). *The many voices of fluency.* Paper presented at the BAAL Annual Meeting.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied linguistics, 17*(1), 84-119.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of memory and language, 34*(6), 709-738.

van der Maas, H. L., & Molenaar, P. C. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological review, 99*(3), 395.

Van Dijk, M., Hunnius, S., & van Geert, P. (2012). The dynamics of feeding during the introduction to solid food. *Infant Behavior and Development, 35*(2), 226-239.

Van Dijk, M., & Van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development: An International Journal of Research and Practice, 16*(1), 7-33.

Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal, 92*(2), 179-199.

van Geert, P. (2019). Dynamic systems, process and development. *Human development, 63*(3-4), 153-179.

van Geert, P., Steenbeek, H., & Kunnen, S. (2012). Monte Carlo techniques: Statistical simulation for developmental data. In *A dynamic systems approach to adolescent development* (pp. 59-68): Psychology Press.

van Geert, P., van der Maas, H., & Savelsbergh, G. J. (1999). Non-linear developmental processes. *Royal Netherlands Academy of Arts and Sciences*.

Van Geert, P., & Van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development, 25*(4), 340-374.

Van Patten, B. (2002). Processing instruction: An update. *Language Learning, 52*(4), 755-803.
Van Patten, B. (2003). *From input to output: A teacher's guide to second language acquisition*: McGraw-Hill.
van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive psychology, 19*(4), 403-440.
Verspoor, M., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development: Methods and techniques* (Vol. 29): John Benjamins Publishing.
Verspoor, M., Lowie, W., Chan, H. P., & Vahtrick, L. (2017). Linguistic complexity in second language development: Variability and variation at advanced stages. *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle, 14*(14-1).
Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal, 92*(2), 214-231.
Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (Vol. 27, pp. 27-62). Amsterdam: John Benjamins.
Wendel, J. N. (1997). *Planning and second-language narrative production.* Temple University,
Wenden, A., & Rubin, J. (1987). *Learning strategies in language learning*: Prentice-Hall International.
Wenden, A. L. (1998). Metacognitive knowledge and language learning1. *Applied linguistics, 19*(4), 515-537.
Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing, 30*(2), 231-252.
Witton-Davies, G. (2014). The study of fluency and its development in monologue and dialogue. *Unpublished doctoral dissertation, Lancaster University, Lancaster, UK*.
Zeeman, E. C. (1976). Catastrophe theory. *Scientific American, 234*(4), 65-83.
Zhang, S. (2009). The Role of Input, Interaction and Output in the Development of Oral Fluency. *English Language Teaching, 2*(4), 91-100.

Appendices

Appendix 1: Questionnaires

Questionnaire about Learning English (Motivation)

☐ I have received information regarding this research and had an opportunity to ask questions. I believe I understand the purpose, extent and possible risks of my involvement in this project and I voluntarily consent to take part.

I'm interested in your personal opinions to help me plan your class. This is NOT a test. The results will NOT be used to give you a grade OR to evaluate your performance. Please answer each question honestly so I can plan a class best suited to you.

Please put an 'X' in the box that best expresses how true the statement is about your feelings or situation.

Example: If you really like watching movies a lot, you should put an 'X' in the first box (Absolutely true)

	Absolutely true	**Mainly true**	**Partly true, partly false**	**Not so true**	**Not true at all**
I like watching movies a lot	X				

Please remember: There are no right or wrong answers. I'm only interested in your personal opinion.

	Absolutely true	**Mainly true**	**Partly true, partly false**	**Not so true**	**Not true at all**
1. I have my own ways of studying English vocabulary					
2. I often chat in English on the Internet.					
3. My friends think that studying English is important.					
4. I study English because I would like to spend some time abroad.					
5. I am certain that I will be able to use English successfully in my future job.					
6. I use English language-teaching computer programs.					
7. Even if I'm well prepared for the class, I feel anxious about it.					
8. I am happy when I see that I am making progress in English.					
9. If I can't understand something in the English class, I ask others (my English teacher, friends etc.) for help.					
10. I am certain that I will be able to write emails and other correspondence in English.					
11. I study English as it is necessary to pass my exams.					
12. Solving a task in English makes me feel good.					
13. Studying English will help me feel part of the					

	Absolutely true	Mainly true	Partly true, partly false	Not so true	Not true at all
international community of people speaking English.					
14. I can imagine myself reading books and articles in English					
15. If there is something that I do not understand in the English class, I check it out on the Internet.					
16. I keep up to date with English by working on it almost every day.					
17. My friends are not bothered to study English.					
18. I often use the Internet to practice English.					
19. I try to find opportunities to read in English.					
20. I have my special techniques to overcome boredom when learning English.					
21. I have my special techniques to make learning English interesting.					
22. I would feel uneasy speaking English with/to a person who spoke that language					
23. People around me tend to think that it's a good thing to know foreign languages.					
24. I feel embarrassed to volunteer answers during the English class.					
25. My parents really encourage me to study English.					
26. Learning English is really great.					
27. The things I want to do in the future require that I speak English.					
28. I am certain that I will be able to read books and newspapers in English.					
29. I'm ready to work hard to learn English.					
30. If I could speak English well, I could get to know more people from other countries. (Not only from English speaking countries)					
31. I worry about the consequences of failing tests, assignments and exams in English.					
32. I am certain that I will be able to get my ideas across clearly when speaking English.					
33. I study English because I'd really like to be good at it.					
34. I can get so nervous during the English class that I forget things I know.					
35. My parents encourage me to practise my English as much as possible.					
36. I really enjoy learning English.					
37. When I imagine my future job, I see myself using English.					
38. When I study English, I often do more than is necessary.					
39. I try to find opportunities to practice English.					
40. Studying English will help me understand people from all around/over the world. (Not only English speaking countries)					
41. I try to prepare for every English lesson, even if I know that I won't be tested.					
42. I feel more tense and nervous in my language class than					

	Absolutely true	Mainly true	Partly true, partly false	Not so true	Not true at all
in my other classes.					
43. My parents consider the subjects of foreign languages important at school.					
44. I get nervous when I'm speaking in my English class.					
45. My parents have emphasised how important English will be for me in the future.					
46. I find that learning English is really interesting.					
47. I need English for my future career.					
48. My friends think English is cool.					
49. I would really like to communicate with foreigners in the future.					
50. I can imagine myself writing emails in English.					
51. I would like to be able to use English to communicate with people from other countries.					
52. I study English because it will be necessary to work worldwide.					
53. I rarely put off my English homework until just before it is due.					
54. I'm afraid that other students will laugh at me when I speak English.					
55. My parents feel that I should really try to learn English.					
56. I am certain that I will be able to get my ideas across when writing in English.					
57. I study English because I'd really like to be good at it.					
58. I plan my preparation and reviews before English tests.					
59. I can imagine myself speaking with high proficiency.					
60. I study English as it will help me to earn a lot of money.					
61. If there is something that I do not understand in the English class, I make efforts to find out more about it.					
62. I am certain that I will be able to make a conversation in English.					
63. I imagine myself speaking English with/to international friends.					
64. Learning English is necessary because it is an international language.					
65. I can honestly say that I'm really doing the best I can (my best) to learn English.					

Your Background* and Input on the Class

Name :
Student Number :
Gender : Female Male
Age :

What is your most recent TOEIC score? Date of Test:
What is your most recent TOEFL score? Date of Test:

	Absolutely true	Mainly true	Partly true, partly false	Not so true	Not true at all

How old were you when you started learning English?
What is your mother's highest level of education?
JHS HS Vocational School University: BA MA PhD
What is your father's highest level of education?
JHS HS Vocational School University: BA MA PhD
Do you have experience travelling or studying abroad? Please explain.
What are your own ideas about how you learn English?
What would you most like to do in this class?

*This information is only for helping me to improve this course. All personal information will be kept confidential.

Self-regulation Questionnaire

Please put an 'X' in the box that best expresses how true the statement is about your feelings or situation.

Example: If you really like skiing a lot, you should put an 'X' in the first box (Absolutely true)

	Absolutely true	Mainly true	Partly true,	Not so true	Not true at all
I like skiing a lot	X				

Please remember: There are no right or wrong answers. I'm only interested in your personal opinion.

Questions	Absolutely true	Mainly true	Partly true	Not so true	Not true at all
1. I am learning English because it is instructed by my department.					
2. I know what is important to learn when studying English.					
3. When studying English, I can concentrate for a long time.					
4. When the study material is difficult, I ask for assistance.					
5. I am learning English because my teachers instruct it.					
6. When studying English, if my friends call me, I give up work and go.					
7. I evaluate my overall English progress.					
8. When studying English, I translate everything					

Questions	Absolutely true	Mainly true	Partly true	Not so true	Not true at all
into Bahasa Indonesia.					
9. I am learning English because my society wants it.					
10. I can provide enough time to study English.					
11. When the study material is difficult, I give up studying.					
12. When studying English, I understand the tasks.					
13. I am learning English because my future job requires it.					
14. I evaluate my exam results.					
15. I can provide enough time to do my homework.					
16. When studying English, I can use my materials efficiently.					
17. I am learning English because I want to communicate with foreigners.					
18. When my English progress drops, I study more.					
19. I am satisfied with my English progress.					
20. When studying English, I find outside school sources to help me.					
21. I am learning English because I want to be successful in life.					
22. I believe I can overcome my learning difficulties.					
23. When the study material is difficult, I skip it and find an easier one.					
24. I am learning English because I want to be successful in my future job.					
25. When studying English, I work on the tasks in order of importance.					
26. I can provide enough time to review on my English exams.					
27. I am learning English because I want to be able to use technology better.					
28. When studying English, I plan my study ahead.					
29. When the study material is difficult, I search for alternative ways to understand and complete it.					
30. I am learning English because I want to go abroad.					

Appendix 2: Task's Description and Materials

Task instructions and analysis

Task 1: Correcting error in a news information video

This task is based on a listening activity (video) which requires students to watch and listen to a news video in order to complete a task. In this task, students are given a fact sheet containing list of words taken from the video (see Material part). Some errors in form of incorrect dates or details are also put in the list. These errors will be the students' task to find and to correct.

	Correcting a news information error and sequence
Gap	Students have a list of words with some of them are incorrect.
Resources	Students have to watch and listen to the video to find and correct the errors.
Outcome	Students find the errors and correct them.
Primary focus of attention	Students focus on identifying the differences between what was said and what was written down based on the information in the video and the list.

Task 2: Retelling a news for a friend

This task is called 'retelling a news for a friend' which is based on Ellis's (2005) 'text reconstruction' task. In this task, two students are given two different news stories that they have to read and summarize. After reading and summarizing, the news stories are removed and the students tell the stories to one another. Each listener is given a fact sheet containing a list of information from the news s/he has read. The listener has to notice the events occur in the story based on the speaker's report. Meanwhile, the speaker has to present the story that contains information listed in listener's fact sheet.

	Retelling a news for a friend
Gap	The listener completes his/her fact sheet based on the speaker's report, while the speaker needs to retell the story to match the information in listener's fact sheet.
Resources	Students process the text to understand the story and present a report for his/her friend.
Outcome	The listener can notice all the information given by the speaker, while the speaker can inform all the required information to the listener.
Primary focus of	Students focus on summarizing and presenting a news information about a terror incident for a friend. They have also to focus on matching the information presented by a friend with the list of information they have.

attention

Task 3: Summarizing a news story for radio audiences

This task requires students to present a 1 to 2-minute summary of a news report based on a news video. First, they watch and listen to a new news video about a terror incident and then they summarize the news story and prepare their own fact sheet that will be used for the report. Students are told that they are going to present the report on air in a radio station in their campus. The audience of the radio station are students, teachers, academic staffs, and the management in the campus.

	Summarizing a news story for radio audiences
Gap	Students need to create a fact sheet that they will use for their news report.
Resources	Students rely on their own linguistic and knowledge background and reflect from previous tasks to create the fact sheet and present a news report.
Outcome	Students can present a news report with all the important information as stated in their fact sheet.
Primary focus of attention	Students focus on presenting the information in the fact sheet into a radio news report.

Task Materials

Week 1 (Pre-test)

Video title : Muslims condemn Jakarta terror attack – 9 News Perth

Video source : https://www.youtube.com/watch?v=_lO23npLS5E&t=44s

Week 2 (Practice 1)

Topic : Terror Incident

Task 1

Video title : Terror in Egypt: Hundreds killed in Sinai mosque attack

Video source : https://www.youtube.com/watch?v=jl66xdkuRrQ

Fact Sheet :

No.	Words	Tick or Cross	Correction
1	Northern Sinai		
2	Deadly		
3	250 people		
4	Inside a mosque		
5	40 Militant		
6	Assisi seize power in 2014		
7	Full power response		
8	Trump tweed support		
9	Decrease in militant activity		
10	Sunni Moslems targeted		

Task 2

Reading 1 title : Egypt hunts for killers after mosque attack leaves at least 235 dead

Sources : http://edition.cnn.com/2017/11/24/africa/egypt-sinai-mosque-attack/index.html.

Fact Sheet :

No	Information	Tick or Cross
1	Military's hunt	
2	Friday assault	
3	brute force	
4	a bomb	
5	The gunmen	
6	ambulances	
7	automatic weapons	
8	Three days of national mourning	
9	Ministry of interior	
10	Sunni Moslems	

Reading 2 title : Egypt launches air strikes against suspected terrorists after mosque terror attack kills 305

Source : https://www.independent.co.uk/news/world/africa/egypt-mosque-terror-attack-latest-airstrikes-witnesses-bir-al-abed-sinai-isis-terrorists-a8075356.html

Fact sheet :

No	Information	Tick or Cross
1	Egypt air force	
2	vehicles linked to the terror attack	
3	Terrorists killed	
4	a mountain 40 km west of Sinai	

5	Egypt's President	
6	Death toll rose 305	
7	ISIS is suspected	
8	Sinai has been targeted since 2011	
9	"Sinai State"	
10	Sufi's shrines in Northern Sinai	

Task 3

Video title : Hostage-taker in France shot and killed

Video source : https://www.youtube.com/watch?v=HtKUvtkAjDE

Week 3 (Practice 2)

Topic : Road Accident

Task 1

Video title : Road Accident, 9 News Perth

Video source : https://www.youtube.com/watch?v=ZwF8lA0m-IU

Fact Sheet :

No	Information	Tick or Cross	Correction
1	The third crash in three weeks		
2	A seventy-six years old woman and her husband		
3	a black Hyundai		
4	RAV4		
5	Kewdale		
6	targeted for the next wave of fixed speed cameras		
7	Road Safety Minister		
8	Forest Highway		
9	a high speed crash near Seabird		
10	Another four people		

Task 2

Reading 1 title : Florida car crash kills four British family members

Sources : https://www.independent.co.uk/news/world/americas/british-family-killed-florida-car-crash-titusville-bristol-kennedy-space-centre-a8287156.html

Fact Sheet :

No	Information	Tick or Cross
1	4 people killed	
2	Mitsubishi Saloon	
3	Back to rental home	
4	Titusville Police Department deputy Chief	
5	A fifth member of the family	
6	Davenport	
7	an illegal U-turn	
8	heavy traffic	
9	Minor injury	
10	UK Foreign Office spokesman	

Reading 2 title : Police arrest teens over deadly hit-and-run crash in Melbourne

Source : http://www.abc.net.au/news/2018-04-21/two-people-dead-after-hit-and-run-involving-stolen-car-wantirna/9683384

Fact sheet :

No	Information	Tick or Cross
1	Two teenagers	
2	A married couple	
3	Two died	
4	Two injured	
5	Knox Club	
6	A jacket	
7	Acting Superintendent Stuart McGregor	
8	The teenagers may have significant injuries	
9	The stolen car	
10	Thieves	

Task 3

Video title : Tragic Accident, 9 News Perth

Video source : https://www.youtube.com/watch?v=Cat6cQ2H73k

Week 4 (Practice 3)

Topic : Natural Disaster

Task 1

Video title : Europe floods: Death toll rises as France declares natural disaster - BBC News

Video source : https://www.youtube.com/watch?v=yHGy3wp61YY

Fact Sheet :

No	Information	Tick or Cross	Correction
1	Surging torrent		
2	Heavy rain		
3	Damage and loss of lives		
4	Flash flood		
5	South of Paris		
6	Evacuated houses		
7	Murky shape of a car		
8	Sense of solidarity		
9	The France Capital		
10	European Football Championship		

Task 2

Reading 1 title : Kashmir avalanche death toll climbs to 24

Sources : https://www.aljazeera.com/news/2017/01/avalanches-kills-10-indian-soldiers-kashmir-170126095124043.html

Fact Sheet :

No	Information	Tick or Cross
1	Rescuers	
2	20 troops killed	
3	army post	
4	Gurez area	
5	Four people from a single family killed	
6	A survivor	
7	2012	
8	one of the most severe winters in recent decades	
9	The timesofindia.com website	
10	Villagers from Waltengoo Nar	

Reading 2 title : High school students hit by avalanche in Tochigi's Nasu

Source : https://www.aljazeera.com/news/2017/03/high-students-hit-tochigi-avalanche-170327035931601.html

Fact sheet :

No	Information	Tick or Cross
1	Seven students and a teacher	
2	More than 100 troops	
3	a three-day mountaineering expedition	
4	Otawara High School	
5	Doctors confirmation	

6	40 people injured	
7	The town of Nasu	
8	a surface avalanche	
9	an annual event	
10	The ski resort had been closed	

Task 3

Video title : Freak hailstorm surprises Russian beach goers – BBC News

Video source : https://www.youtube.com/watch?v=RZ2XgZ-6NMo

Week 5 (Practice 4)

Topic : Airplane Accident

Task 1

Video title : Taiwan Plane Crash – Rescue and recovery after TransAsia plane crash - CNN

Video source : https://www.youtube.com/watch?v=yHGy3wp61YY

Fact Sheet :

No	Information	Tick or Cross	Correction
1	Footage		
2	CNN's affiliates		
3	Pitching		
4	68 passengers		
5	There were no survivals		
6	100 first responder		
7	Recently serviced		
8	Distraught		
9	Their attitudes is good		
10	Flight data recorders		

Task 2

Reading 1 title : Iran plane crash: 65 killed as commercial flight goes down in southern Iran

Sources : https://www.cbsnews.com/news/iran-plane-crash-66-killed-as-commercial-flight-goes-down-in-southern-iran/

Fact Sheet :

No	Information	Tick or Cross
1	A commercial airplane crashed in southern Iran	
2	foggy and mountainous area	

3	65 people on board killed	
4	59 passengers and 6 crews	
5	Rescue helicopters	
6	Mount Dena	
7	The Aseman airline banned	
8	FlightRadar 24	
9	1994 and 2008 accidents	
10	Iranian Red Crescent	

Reading 2 title : Aylesbury mid-air crash: Four dead as plane and helicopter wreckage lands near Rothschild manor house

Source : https://www.telegraph.co.uk/news/2017/11/17/aylesbury-mid-air-crash-fatalities-feared-afteraircraft-helicopter/

Fact sheet :

No	Information	Tick or Cross
1	A helicopter and a light aircraft collide	
2	no air traffic control	
3	Staff shortage	
4	2 people in each aircraft	
5	4 people killed	
6	The wreckage	
7	Loud bang and plume of smoke	
8	Off duty fire fighter	
9	emergency services	
10	7 fire appliances	

Task 3

Video title : Nepal plane crashes minutes after take-off

Video source : https://www.youtube.com/watch?v=hPWLCKBSwBw

Week 6 (Practice 5)

Topic : Forest Fire

Task 1

Video title : Portugal forest fire – At least 61 dead in nation's worst fire in decades I NBC Night Sky News

Video source : https://www.youtube.com/watch?v=F4F4BU-wYvs

Fact Sheet :

No	Information	Tick or Cross	Correction
1	A deadly inferno		
2	A woman inside her house		
3	entire villagers were evacuated		
4	Filipa Soares		
5	Portugal's latest heatwave		
6	Pedrogao Grande		
7	700 firefighters		
8	Spain and France		
9	Three days of mourning		
10	Portugal's national soccer team		

Task 2

Reading 1 title : Oklahoma wildfires kill 1, force evacuation

Sources : https://edition.cnn.com/2018/04/13/us/oklahoma-wildfires/index.html

Fact Sheet :

No	Information	Tick or Cross
1	a 61-year-old man died	
2	A 54-year-old hunter	
3	1,000 people evacuated	
4	the National Weather Service	
5	Woodward County	
6	200,000 acres	
7	dry conditions and strong winds	
8	additional evacuations	
9	State Highway 34	
10	a trio of dangerous weather	

Reading 2 title : Wildfires hit French Riviera, thousands evacuated

Source : https://edition.cnn.com/2017/07/26/europe/france-wildfire-evacuations/index.html

Fact sheet :

No	Information	Tick or Cross
1	More than 10,000 residents and tourists	
2	a beach near Saint-Tropez	
3	A combination of strong winds, high temperatures, and a lack of rain	
4	Paul Adriaansens	
5	Over 100 firefighting operations	

6	Around 3,000 campers	
7	Matthieu Dany	
8	Londes-les-Maures	
9	more than 4,000 firefighters and soldiers	
10	President Emmanuel Macron	

Task 3

Video title : Indonesia forest fires smoke chokes Singapore

Video source : https://www.youtube.com/watch?v=th_MX9wgtu8

Week 7 (Practice 6)

Topic : Accident on the Sea

Task 1

Video title : Cruise ship sinking in Italy: 6 bodies found

Video source : https://www.youtube.com/watch?v=_rjy9wek_yg

Fact Sheet :

No	Information	Tick or Cross	Correction
1	Six victim		
2	A significant error		
3	60 people missing		
4	72 hours searching		
5	coastguard		
6	already submerged		
7	central figure		
8	The Captain is being blamed		
9	Serious error of judgement		
10	He has been punished		

Task 2

Reading 1 title : Tanker and ship collision near Shanghai leaves 32 missing

Sources **http://www.bbc.com/news/world-asia-china-42594249.**

Fact Sheet :

No	Information	Tick or Cross
1	a cargo ship	
2	Near Shanghai	
3	Joint-rescued	

4	32 people	
5	from Iran to South Korea	
6	136.000 tonnes of oil	
7	Floating and burning	
8	A big spill	
9	keep the oil offshore	
10	21 rescued	

Reading 2 title : US Navy ship and oil tanker collide near Singapore

Source : **http://www.bbc.com/news/world-asia-40995829**

Fact sheet :

No	Information	Tick or Cross
1	US destroyer	
2	Near Singapore	
3	21 August 2017	
4	10 sailors are missing, 5 injured	
5	a routine stop	
6	The second collision	
7	Damage on the left-hand side	
8	Damage to the front tank	
9	No tanker's sailors injured	
10	No oil spill	

Task 3

Video title : Cruise ship accident leaves crew fatalities

Video source : https://www.youtube.com/watch?v=6vS9JwoAru4

Week 8 (Practice 7)

Topic : Domestic Violence

Task 1

Video title : Men suffer domestic violence too

Video source : https://www.youtube.com/watch?v=ZGDTDawB4wE

Fact Sheet :

No	Information	Tick or Cross	Correction
1	New Campaign		
2	A happy man		
3	Mildly violence		
4	Steam iron on right arm		
5	Lost 25 % of bodyweight		

6	Left shoulder		
7	80 years in prison		
8	Rebuilding body and life		
9	Gender neutral crime		
10	No help for male victims		

Task 2

Reading 1 title : Domestic violence abuser called police to lock up his partner

Sources https://www.canberratimes.com.au/national/act/domestic-violence-abuser-called-police-to-lock-up-his-partner-20180919-p504q4.html

Fact Sheet :

No	Information	Tick or Cross
1	A Canberra man	
2	Pleaded guilty	
3	reconciled	
4	10 days in custody	
5	the violence continued	
6	mid-January	
7	a statement to police	
8	a limited criminal history	
9	four years jail	
10	a significant impact upon the victim	

Reading 2 title : Ugly spike in deaths related to WA domestic violence

Source : https://www.news.com.au/national/western-australia/ugly-spike-in-deaths-related-to-wa-domestic-violence/news-story/9e27ee9ebf8ed86270e31f621f07a08f

Fact sheet :

No	Information	Tick or Cross
1	23 murdered	
2	The horrific murders in Bedford	
3	Carlisle mother	
4	This year's alarming spike	
5	Simone McGurk	
6	Hide it well in public	
7	Scared to speak up about	
8	Kedy Kristal	
9	284 women and six children	
10	The State Government	

Task 3

Video title : Former Memphis news anchor shares her story of surviving domestic violence

Video source : https://www.youtube.com/watch?v=eKM6_Gn3ewQ

Week 9 (Practice 8)

Topic : Hate Crime

Task 1

Video title : Kansas City triple shooting investigated as hate crime by FBI

Video source : https://www.youtube.com/watch?v=52J_jrJHauk

Fact Sheet :

No	Information	Tick or Cross	Correction
1	A sport bar		
2	A college baseball game		
3	51 years old killer		
4	Around 7.50 pm		
5	The killer was from same town		
6	The killer was a former military		
7	One customer called police		
8	Two Indian-decent victims		
9	Two victims died		
10	Ian Grillot was shot at his stomach		

Task 2

Reading 1 title : Police investigate 'hate crime' against Indian family

Sources : https://www.thehindu.com/news/international/police-investigate-hate-crime-against-indian-family/article24988419.ece

Fact Sheet :

No	Information	Tick or Cross
1	an Indian family	
2	The Metropolitan Police	
3	a large fire at his home	
4	a group of four to five young people	
5	managed to escape	
6	voluntary unpaid work	

7	critical of the police response	
8	rising hate crimes.	
9	prejudice	
10	increase in 2019	

Reading 2 title : Hate Crime at Austin Mosque?

Source : https://www.austinchronicle.com/news/2018-09-21/hate-crime-at-austin-mosque/

Fact sheet :

No	Information	Tick or Cross
1	the target of vandalism	
2	this level of vandalism	
3	Middle Eastern grocery store	
4	The APD's request	
5	the same individual	
6	83% nationally	
7	anti-Muslim hate groups	
8	extra security precautions	
9	The Imam's first concern	
10	working with APD	

Task 3

Video title : Hate-motivated crimes in Canada

Video source : https://www.youtube.com/watch?v=8VNdG3ygM3A

Week 10 (Post-test)

Video title : Indonesia Earthquake: Tsunami Warning Triggers Panic – Al Jazeera

Video source : https://www.youtube.com/watch?v=ag-SCzzxIT0

Appendix 3: News Video Transcripts

(Pre-test, Post-tests and task performances)

Week 1 (Pre-test)

At the capital main mosque, prayer for those killed and injured, and fought on destructive violent extremists like Isis, recruiting Indonesian in the name of Islam.

“In a strongly worded sermon during Friday prayer here at the grand mosque in Jakarta, the Imam has said that terrorism is a crime against humanity, and that Islam shouldn’t be something to be scared off or used to spread fear among people. This, the first attack on Indonesian soil, claimed by Isis, has compelled people here to speak out against terrorism in the name of Islam.”

“What happened yesterday, they are not Muslims, they are infidel. I condemned them. I demand the government solve this problem. As a Muslim, I reject this terrorism in the name of Islam.”

‘Islam becomes the escape goat. People believe Islam is identical to violence. Every time violence happens, people think that it’s related to Islam, but it’s not.”

“no no no, I’m not terrorist and then Muslim in Indonesia not terrorist. But I want to say it to the world that human in the world err Muslim, Islam, Muslim not terrorist or terrorist not Muslim.”

Across the city, Nahdatul Ulama, an association of Islamic scholars with fifty millions Indonesian members worldwide held an inter-faith dialogue with representative of Catholic and protestant churches, Jewish, Hindu, and Buddhist community, calling for unity and peace.

In November 2015 recognizing the potential threat, the association released a video campaign, aimed at countering the extremist ideology. There’s growing concern heighten by this attack, about Isis gaining a foothold in the world largest Muslim country. At least 500 fighters are believed to have travelled to Iraq and Syria from here. But that 500 hundred out of 200 million moderate Muslims in a secular state.

Isis claims responsibility for Thursday gun and bomb attack saying that the group targeting the crusaders alliance which is fighting the Islamic state. “Not in our name”, said the Indonesian Muslim.

Sahima Mohsin, cnn Indonesia, Jakarta.

Week 2

In just the last couple of minutes, it looks like this very nasty situation in the south-west of France has been resolved. According to report, the attacker who has been, quote-unquote,

neutralized, we believe killed. It all started in the city of Carcassonne let again in south-west France. The suspect shot at four police officers, badly wounded one, and then shot of a car that he needed for getting away and killed one person there, then drove over to a nearby village, entered a supermarket, took hostages and screaming out; according to eyewitnesses "Allahu Akbar" God is great and claiming that he was representing ISIS and demanding that a terrorist suspect now on trial related to a terrible terrorist attack back to 2015 in Paris be released.

In that supermarket he shot dead we believe two other people. There was a stand-off between the police and the gunmen we believe to be Moroccan origin. The police offered him a release that is to give himself up. His family was there, but he did not. And literally just 15 minutes ago the police assaulted the supermarket. They shot and neutralized the suspect. We believe he has in fact been killed.

Off course we all know that France has seen a lot of terror in the past few years. Two major attacks in Paris in 2015 and the horrendous attack in Nice in 2016. This area of south-west France has been attacked as well. The Prime Minister and the President said that they believe that this is a terror attack and their major official are already at the scene, and they treated this as ISIS, at least, inspired that has led several dead, several wounded.

Week 3

Good evening. It's been a while and tragic day on our road leaving two people dead. The details are just emerging but what we know so far is a pair was killed in a freak accident 200 kilometres North of Perth.

This is a view from a 9 News helicopter a short time ago. The early indications are that a truck has lost control and crossed to the wrong side of the road, and hit a car head on.

Scott Cunningham is tracking the development. Scott, early stages of the investigation but what do we know so far?

Michel, it is a particularly tragic case. What we've been told is that the truck has a tyre blow up before it lost control and crossed onto the wrong side of the road. As you can see from the helicopter's footage you can see that the car was ripped apart and then caught fire leaving the occupants no real chance at survival.

We have been spoken to the WA Police tonight, and they told us the victims are yet to be formally identified.

The out of control truck then ploughed into the bushland. We are told that the driver suffered a minor injury.

As we go to air tonight that section of Brent Highway about 15 Ks South of Badgingarra remains close to traffic. Diversion has been put in place as major crash officers investigate, Michel.

Scott, thank you for the update.

Week 4

The beach goers have had been a pleasant time in summer sunshine, then suddenly everything changed. The sky opened and a freak storm halted the sun worshippers with giant hail stones.

Those in the sea run to the shore. Many clearly intend from the fracturing they were receiving. Others thought what cover they could, some haggled together on the beach umbrellas which offered little in the way of protection.

It was another 15 minutes before the storm has eventually subsided. Russia has been experiencing extreme weather all summer. Many central areas have snow and freezing temperature.

At the Lake Side resort not far from Nova suburb two girls died when a tree fell on their tent as they sheltered from the latest storm. The authorities have been an investigation. John Breen, BBC News.

Week 5

This is the scene after a Sita Air plane crashed just outside Nepal capital in clear weather. The twin-engine propeller-driven Dornier aircraft was flying to Lukla, a small town frequented by hikers heading to Mount Everest. It plunged the ground less than a kilometre away from Kathmandu International Airport only minutes after take-off.

“It carried 16 passengers and three crew members, all confirmed dead. All dead bodies have been collected.”

A national police spokesman said the plane came down near a river and that the pilot seemed to have tried to land safely on the bank of the river. But unfortunately, the plane caught on fire. The flame quickly engulfed the plane. A witness heard screaming from inside the aircraft. She said people who’d come to help, couldn’t get close enough to throw water on the plane because they were scared it was about to explode. Even when the firemen have arrived, it took some time to put up the blaze.

Thousands of tourists head to the Himalayas each year, especially in autumn, the peak tracking season. Because of the poor road infrastructure, many people fly to start of their height using one of the 16 domestic airlines and 49 airports to reach the remote areas. But this is the sixth fatal air crash in Nepal in less than 2 years. Only 4 months ago 14 people died when an Agni airplane carrying tourists crashed in its way to Nepal's Annapurna mountain range. And a year ago a small plane taking tourists to Everest hit a hill side near Kathmandu.

Despite this latest crash, thousands of visitors continue to use this small plane. Even as Nepal’s air safety record gets worst.

Week 6

Singapore's usually clear sky are choked with smoky haze, as a part of Malaysia. And a growing number of people are complaining of respiratory problem.

"It's very bad this year. And I have difficulty breathing when I come out, so… the mask at least we have the mask but it is hard a bit."

A local leader watches a smoke from a forest fire in Sumatera bellowed into the sky. It's an annual event but he says some fire like this one should never have been late.

"They never inform us about this activity. I checked with the office from forest ministry, they have no permission. So this is illegal."

The challenging terrain of Sumatera's forestall areas means that some of the fires are hard to locate, and even harder to monitor. Indonesia's neighbours wanted to take strong action so the region can avoid air travel and business disruption, and further harms to people's health.

Seize of no fire draw attention to the state of Indonesia forest. Earlier this year, the government extended the moratorium on new logging licenses by two more years. The aim is to protect the archipelago remaining primary forest, and control land clearing.

Nidhi Dutt, Aljazeera, Jakarta.

Week 7

It was a safety drill that turned deadly. Crew members aboard the Thompson Majesty Cruise ship was performing a mock rescue when the cables of the lifeboat they were in snapped, sending the craft plummeting twenty meters into the ocean. Rescue personnel recalled to the port after report that several people had gone overboard. The cruise company issued a statement saying:

"Thomson Cruises is aware of an incident involving the ship's crew on board Thomson Majesty in La Palma, Canary Island. We are working closely with the ship owners and managers to determine exactly what has happened and provide assistance to those affected by the incident."

Nearly 1500 passengers were on board of the ship at the time. They had been due to sail to Portugal on Sunday. Spanish officials have not released the identities of the victims, but they were all crew members. Three others suffered non-life threatening injuries.

Katherine Stencil, Aljazeera.

Week 8

News at 4.30

She was a well-known face in Memphis although she does not live in the love city anymore. Dee Griffin showed ourselves that some people had not seen before today. She is a domestic violent survivor. Griffin spoke at a WCA benefit luncheon where proceed to go toward helping victims. Our reporter talks with Griffin who gives an urgent plea to women to get out of this situation.

Former Memphis anchor Dee Griffin didn't think domestic violent could happen to her until what started as a fairy tale marriage at 2010 ended with her fighting for her life.

"We worked through it. The baby was born in May of 2011 and we were very happy. But then during the fifth week when my son was five weeks old he snapped. My ex-husband snapped. And he physically abused me, hand s around my neck."

It only got worse for Griffin. She says a neighbour who heard her screaming called police.

"I see the hate in his eyes. And someone pinning you up against refrigerator and cabinet and squeezing your neck where you couldn't breathe. It was devastating."

Griffin made it out of the marriage and says her son gave her strength to live.

Some of the people we spoke with who attended this luncheon inside this room said it really encouraging to hear from a former Memphis anchor.

"It makes me feel enlighten. And the reason why I say that is because a lot of time we have the misconception that this only happen to maybe certain people. People who maybe below the poverty line."

Makita Odom said that hearing from someone like Griffin will encourage other women to walk away.

"This could happen to anyone. She was a professional who had everything go on in life. And it happened to her."

"Get out. Save yourself. Save your family!"

Safin Riley reporting.

Week 9

For the second time in a week, a video of a racially charge confrontation in Ontario has been posted online. That's let people to question whether the authority should treating these cases as hate crimes. This most recent incident happening outside a Walmart near Hamilton after a dispute over a parking space. The video begins with the man on the left confronting the driver. The man's wife is taking the video and it appears to get hit by the truck. It ended with this:

“I don’t like you. I don’t like her. I will kill your children first.”

The driver has been charged with threatening death, dangerous driving, and failing to remain at the scene of an accident. And police said that they are reviewing it as a possible hate crime.

This follows another encounter caught on video. This one is in Toronto. The man with the red shirt has been charged with two kind of assaults and one kind of a threatening death. The Toronto’s police say they do consider this as a hate crime. This video has led two a vigorous debate on social media. Among the issue: were they hate crimes? And if so, what the consequences be?

A legal expert we talked to said the words and actions in the video appear to him to be motivated by hate. And if the judge agrees, that would make jail time likely.

“Hate is a motivation for a crime constitute a strong aggravating feature which is court should take into consideration and should impose a deterrence sentence. It has to send an important message to the members of the community who feel marginalised that they will be protected, and it has to send important message to those who might be like-minded that hate comes with a consequence.”

Sandler pointed out that maybe more to what happen than we can see in the videos. But those videos also make the conducts of both men very clear.

Week 10 (Post-test)

As the earthquake strike, people in Banda Aceh fled from buildings and spilled onto the streets. The earth shook for over a minute. Then shock turned to panic that a tsunami might strike, prompting a rush to get to a higher ground. The streets became clogged with cars and people.

The US geological survey measured the quake at 8.6 magnitudes. It hit 20 km below the ocean floor.

Indonesian President Susilo Bambang Yudhoyono tried to reassure his citizens. "The situation is under control. There was a bit of panic, but the people in Aceh have been moved to safety. Our early warning system has worked well. So far there's no threat of a tsunami, but our team remains on standby. I have ordered the National Agency for Disaster Management to fly to Aceh immediately to make sure everything is under control".

Strong aftershocks continued throughout the region, and people remained on alert. It triggered memory of the devastating 2004 earthquake and tsunami that killed 230.000 people. Three quarter (3/4) of them in Aceh. This time, hospitals treated injuries but there were no reports of deaths.

Tsunami disaster alert systems put in place after the 2004 disaster swung into action and warnings were issued across the Pacific Rim. In India building were evacuated and tsunami alert issued in the southern states. Thailand placed six provinces on alert. Buildings were evacuated as a precaution and people moved to a higher ground. On holiday island of

Phuket patients were moved from hospitals. Although this was not the disaster feared, it was proved that the warning system worked.

Ella Callen, Al-Jazeera, Bangkok.

Appendix 4: Students' Transcripts for the Phase Transition Analysis

Week 1

1. Student18

OK good morning ladies and gentlemen. Today, I will to inform the news about Islam re er terrorist. Some to say no no I'm not terrorist. Er Islam is Muslim and then Islam not terrorist. What happen yesterday occur explosion and place crowded and then terrorist still and the panic.

2. Student20

Ok guys, here I'm reporting from Indonesia tragedy. It is about bomb attack in Indonesia. It is the first bomb attack by terrorist that er the suspected by the er Muslim. but there are some of Muslim said and rejected that er there is no way there is no reason Muslim said be a terrorist and terrorist also doesn't mean Muslim. So er also there's some of Muslim said that they have to believe with the government to reduce the bomb the terrorist and about the bomb attack. And that is er also Nahdatul Ulama has said that they are have er fourty million er of Muslim that er recording to the data. Thank you.

3. Student8

Good morning people. This is Zikra Darni reporting from the location. You know that Indonesia is a country that always get terror. And we also know that Indonesia is the Muslim country. And as the narasumber said, I rejected the terrorist in the name of Islam. Islam is not terrorist, Muslim is not terrorist, and the terrorist is not Muslim. Thank you.

4. Student22

Good morning audience. To the news about Jakarta tero terror. The accident is so large and the Islam is something be caught because people call it is very different from from Islam. All the people said that it's from Islamic centre. People say that Muslim in Indonesia isn't terrorist and terrorist is not Islam. And in the twenty fourteen there is happen the same accident or the same problem and it Islamic state. So er so people er judge Islamic community is er the terro terrorist. Thank you.

Week 2

1. Student18

Ok. Good morning everyone. Okay, now I will to inform about gunman claiming link to Islamic state. Suspected Isis terrorist attack supermarket in southern. Police down injured to occur in location. Other people in take with the aim to the location of save. Many people feel fear, so that the police difficult to recur. The incident resulted in one people police suffer injury. The all people in area supermarket feel panic. Okay, thank you very much

2. Student20

Here I'm reporting from the accident in the southern, France. Exactly located in a supermarket. As the suspected with a shot for policemen and one person killed there. The suspected also Isis terrorist in the supermarket. And we know that a few years ago er France also have two attack in the two thousand fifteen and also two attack bombers on the two thousand sixteen.

3. Student8

Good morning everyone. Zikra Darni reporting from the location. There is a nasty condition in in France. The gunman claiming link to Islam in southern France. This bombing in supermarket southern in France. That bomber take hostage in France. The tragedy killing one people, one police office and two people between the supermarket. Zikra Darni reporting from the location. And back to the studio.

4. Student22

Good morning audience. Gunman claiming link to Islamic curt state suspect take in France. Suspected ISIS terrorist takes in a supermarket in the Southern. The gunman claimed holding hostage in France. One person there killed near the supermarket err and shot dead two other people as appear.

Week 3

1. Student18

Ok. Good morning everyone. Meet again, I'm Rina Puji Astuti. I will to inform about traffic tragedy. On the location butchi butchingara. Well, this was a tragic accident and the location was two kilometre near Pearl palace, pearl palace. This today the helicopter saw that a truck lost control. And then there many people are injured at incident there. Thank you.

Okay, thank you very much.

2. Student20

The news has been updated about the crash on the Brand highway, two hundred kilometre in Perth. A truck lost control. This news has been updated by the reporter. Before the truck lost control the tyre of truck was blow up. And two people was killed here. Also the truck crashed along the road. Emm well I am sorry but the truck crossed along the road err and well actually the truck err crashed the car. And the people two people killed right there.

3. Student8

Good morning people. This is Zikra Darni reporting from the location. Today, there is a tragic way of the highway tragedy. The car has been crashed because lost control of the driver and crushed into the wrong side. The truck had tyre blow on Brand highway. Er the tragedy happen five kilometres from the highway. Er because of the tragedy two people were killed and police can identify victims. Thank you for listen. Back to the studio.

4. Student22

Good morning audience. It's the news about highway horror. There happened some accident. The tragedy highway made two people dead. And then the tragic accident there is two people killed on Bucthingara. The truck have a tyre blow on Brandt Highway cause major injury. It happen because a lost control driver that may involve er such many injury. Thank you.

Week 4

1. Student18

Ok. Good morning people. I'm Rina from Queen News channel. I will to inform about hail storm on the beach in America. From this cam it was seen that everyone was very panic running around. The hale of now falling. This happen for fifteen minute and the result in two woman died from fallen tree. Okay Thank you. I'm Rina back to studio.

2. Student20

Here I am reporting from Russia. There at a sea shore directly I am reporting the strangest thing I have ever seen as a reporter. The in a bright summer at a sea shore the people got their holiday enjoy the vocation and then suddenly the weather change become a hail storm. It is a strange thing. This happen around fifteen minutes. And the temperature increase suddenly. Also the people got panic they also protect themselves directly under the bridge umbrella. And then all people wondering and feel weird while when they see directly the sea err you know see directly the the strange event in this year. I am Shiva, reporting from the sea shore Russian.

3. Student8

Good morning everyone. Zikra Darni reporting from the location. A hail storm was happen in the beach of Russia. The wind went up and the temperature suddenly drop from forty one to twenty one degrees. The hail storm was happen for fifteen minutes. Because of this incident two young children was killed and people can be hear shouting at the restless storm seems never ending but there were no report injuries on the beach. Actually, weather has been unusually extreme across Russian this summer with no hitting part of central Russia that rarely see freezing temperature in summer. Thank you.

4. Student22

Good morning audience. Here this event create has happen an hail storm happen hail storm that very unpredicted on the change. When it comes everything change. People was surprised o sustained the expedition yet the expedition change to the show. This is fifteen minutes freezing shock until the beach occur. Thank you.

Week 5

1. Student18

Okay, good morning viewers. Today on Thursday thirteen December two thousand and eighteen, you are still with me. I'm Rina Puji Astuti from Queen tv. I will to inform about Nepal plane crash. The first ever the fire that happen explosion. Nineteen killed all people feel panic. And national police provide assistance for this incident. A lot of safety coming down to help. With help of them resolved. Okay viewer, that is the news today. I'm Rina Puji Astuti, back to studio.

2. Student20

Here I am Shiva, Now reporting about the Nepal air plane crash. This is air plane err kill all the climbers. And well, the climbers are going to the Everest Mountain. There are sixteen passengers that have been killed. And also the air plane have been crashed after fifteen minutes take off.

3. Student8

Good morning everyone. Zikra Darni will report about Nepal airline crashed. Nepal airline crashed on the bank if the river and explode. Seven Britain were among nineteen people killed when a small plane crashed soon after take-off from Nepal airport. Four other tourist believed from China also died along with four Nepal passengers and the plan crews. The flight destination was the small was the small place of Lukla which is the start of many tracts in the Everest region. That's all my report. Back to the studio. Thank you.

4. Student22

Good morning audience. We come with the news. There is Nepal plane crash. It is can seeing in the land city in the shores. Two thousand visitors be in the land. The plane was less more than three years. It's ninety people on the board killed on the Himalaya each year. The plane just only a minute in the crash on the Himalaya each year. Thank you.

Week 6

1. Student18

Ok good morning people. I'm Rina Puji Astuti from Queen Channel. Today I would to re..porting about wild fire forest in Jakarta West Sumatera on June eighteen two thousand and eighteen. So intake in Singapore. And then Singapore to complain at to Sumatera. And now not solution about that. Okay thank you.

2. Student20

Wild fire hit in Sumatera island Indonesia on June eighteen two thousand thirteen. The fire quickly spread at places claim of smoke to various place. Looks very clear the mist over the Sumatera forest. This event cause disturb breathing, shortness of breath and coughing because by the smoke to people. As response, every people must use mask to protect the

breathing. Some foreigner from Singapore and Hindi use mask to protect their breathing also. Okay, thank you.

3. Student8

Good morning people. Zikra Darni reporting from the location. Smoke polluting the sky over Singapore. Smoke from Indonesian fire at unhealthy level in Singapore as authorities push to hand offender. Air pollution in Singapore has risen has risen to the unhealthy level as a quick smoke rise over the island from fire in Indonesia, Sumatera island. Pollution level in neighbouring Malaysia were normal on Friday. A forest complainer for the environmental globe was struggling to enforce law to prevent the drainage of island for plantation and the setting of fire to clear land. Thank you.

4. Student22

There is fire news. It happens in Indonesian. The smoke bellow into the sky. Smoke polluting the other country. It should never concern this. They're just looking at the business not the population or people heart. Many forest land fire and this make bad effect. It is polluting the sky until many location and enough damage other one.

Week 7

1. Student18

Ok good morning people. I'm Rina Puji Astuti from Queen Channel. I will to inform about cruise ship incident. Cruise ship from Spain headed Portugal. They carrying one thousand five hundred passenger. The cause the cause err big ship er up making and then three person felt the three person felt injury. Okay thank you.

2. Student20

I'm reporting now the cruise ship accident has been happened. The Thompson cruises will go to er Spain to Portugal and before that before that travel before that trip the cruise ship err and the crews member will be rescue will be held the rescue the rescue personal training. This is er this has been happened err before they go to the Spain err I'm sorry, they go to the Portugal, and this er when they have the rescue personal training when the crews personal member err go to the life boat, err but unfortunately the cable of life boat was err was broke and err the lifeboat fall down and also err the people who top on them. So err in the cruise on the cruise ship there are fifteen hundred passengers and there are three crews member injured. Err this is happened err this is happen the life boat fall down err un until twenty meter into the ocean. Thank you.

3. Student8

Good morning people. I'm Zikra Darni will report to you about cruise ship accident. The accident of Thomson Cruise er it's happen because safety drill. Crew member were killed in Canary Island. Six crew members have been killed and a quarter three injured after a life boat. That's all my report. Thank you.

4. Student22

Good morning people. There are accident, cruise ship accident. The crew members killed in Canary Island. Thomson is away incident involving the ship's crew and the owner exactly which this happen the more people on the accident. About twenty metre killed incident before have key fault the manager there are fifteen there are fifteen thou there are fifteen hundred missing in the ship. Thank you.

Week 8

1. Student18

Okay good morning. I'm Rina Puji Astuti from Queen Channel. I will to inform is about the domestic violent in University of Memphis. This incident happened on twenty thousand twelve. She was hit by her husband at their first year of marriage. They were fine but think change for an instance. His several tour for several year. He was very lucky because he was saved by his neighbour. And now he was in the field of motivation about the domestic violent. Okay, I'm Rina Puji Astuti. Back to studio.

2. Student20

I'm Shiva and now I'm telling you about the former anchor at the Memphis. She is a violent survivor in the case of domestic and family violent in that city. It began when she got married in two thousand and ten, and the rest of her married with her husband only a year until she had a son. When her son in five week her husband changing directly into a rough person. He did bad thing with on her and fortunately the her neighbour herd her screaming at the time and and err and directly called the police. And now she help another victims of violent and gave her encourage speech err with the with err the others. She said that, save your family, save your children and become be brave and speak up with er your err personality with your experiencing about the violent. Thank you.

3. Student8

Good morning people. This is Zikra Darni will report to you about domestic violence survivor. A former Memphis news anchor who survived domestic violence is sharing her story. She didn't think domestic violence could happen to her until what a start as a fairy tale marriage in two thousand and ten and with her fighting for her life. It happen when her baby was born in May of two thousand and eleven but then during of the fifth week, when their son was five weeks old, he snapped. Her husband, her ex-husband snapped. And he physically abused her, hands around her neck. She said, a neighbour who heard her screaming called a police. She said her husband pinning her up against the refrigerator and squeezing her neck were she couldn't breathe. And then she made it out of her marriage and said her son gave her strength to live. She said that this could happen to anyone. She was a professional who had everything going on in life and it happen to her. She said get out, save yourself, save your family. That's all my report. Thank you.

4. Student22

Good morning people. I'm Sandra Handayani Putri, will present news about domestic violence. The woman said that her husband was very good at first before marriage. But after six, but after five months having a baby, her husband snapped and turning into a violent one. The woman usually hidden and scream, but fortunately saved by the neighbour from the violence in here in her household. So, she can save herself since that. Actually in two thousand eleven the incident have made people felt delay a lot of time. That's all news today. Thank you.

Week 9

1. Student18

Ok good morning everyone. I'm Rina Puji Astuti from Queen Channel. I will to inform about hate crime. This incident happened between tourist and Indian Muslim. The first, tourist crashing Indian Muslim and then here there is a fight of most. Well, atmosphere to be more chaotic. This incident after occur for and reported to the authority. The perpetrator can be sentence to life imprisonment. Well, this incident until now is not finished. Okay, I'm Rina Puji Astuti. Back to studio.

2. Student20

I'm Shiva, I'm reporting about hate crime right now. Nowadays hate crimes widespread into our society and it becomes our concern. And today I will tell you there are two cases about hate crimes. The first case is was happened I'm sorry was happened in err parking space and then err the man and his wife err get err become a victim err by the driver. The driver was a citizen in the city. And err also there are a fight about the parking space. The driver, the real citizen talk and speech err about the hate about the treat them. And then, err his wife the man's wife record the video. And this cases the first cases is err is now the concern by the judge. And the judge will decide it a hate crime err it is a hate crime or not. And the second cases is err is happened I'm sorry was happened at the at the front of some building at a city. And err the real citizen err the resident or the people who grow up in that city err talk err hate speech also with the Muslim with Muslim citizen right there. And the judge err talked that is a hate crimes, hate criminals. And also there is err the and also there is err crime defence lawyer, Mark Sanders said that the hate is a motivation to do the crime and we will have to concern about it and aware about the hate err crime because it is really dangerous for us and for our society. And for punishment of the hate crime is a lifetime. Thank you.

3. Student8

Good morning people. I will report to you about the case of hate crime. The hate crime incident happening outside Walmart near Hamilton over a parking space. Begin with the man on the left confronting the driver. The man's wife is taking the video appears to get hit by the truck. The police said, they are reviewing it as possible hate crime. And the case of hate crime also happen in Toronto. Toronto police say they do consider this a hate crime.

The video was show a debate on social media among the issue, were they hate crime, and if so what would the consequences be a legal expert. That's all my report. Thank you.

4. Student22

Good morning people. I'm Sandra Handayani Putri will present news about hate crime. It is in Toronto, Canada. That is near the hospital one of possible hate crime. He didn't expect this while taking the video that viral. The video was upload in social media, so people know. In this era, everything you do involving picture or video in public are seen. And it cannot hidden by violence alone. That's the news for today. Thank you.

Week 10

1. Student18

Ok good morning everyone. I'm Rina Puji Astuti from Queen TV. I will to inform about Indonesia earthquake. Tsunami warning triggers panic. This happen in Aceh, West Sumatera. The threat of tsunami hitting Aceh has triggered desperate panic. With many coastal resident evacuated their home and plan to higher ground. The earthquake the earthquake in this incident was felt until India and Thailand. The resident trauma because of the incident at twenty thousand four and cause very strong panic. Ok, I'm Rina Puji Astuti, back to studio.

2. Student20

I'm Siva Nasila reporting about earthquake in Indonesia. The earthquake strike Aceh. Right there people right there got panic because the tsunami issue. And the peo there were move to the safety place to the higher ground. And also there is USGS the United State Geological Survey they said that the earthquake has eight point six magnitude. There is no victims right now and there some people injured but there is no right amount about that. And also the people remind about the earthquake two thousand and four. There are two hundred thirty thousand victims er right there, and also this earthquake at the two thousand and two thousand and twelve, there is no er there is no tsunami. The president of Indonesia, Susilo Bambang Yudhoyono said there is the situation is under control and there are and the people er aceh the people in the Aceh got safety place. Thank you.

3. Student8

Good morning people. I'm Zikra Darni will report to you about Indonesia earthquake. The earthquake strike aceh and people get panic about the tsunami issue. And they evacuate to several place and high place. The United States geological survey said that the earthquake has eight point six magnitude. There is no victim right now and some people got injured. The president of Indonesia Susilo Bambang Yudhoyono said that the situation is under control and the people in Aceh get to safety place. That's all my report. Thank you.

4. Student22

Good morning people, I'm Sandra Handayani Putri will present the news for today. Indonesian earthquake in Banda Aceh. It is a big disaster. Indonesian president Susilo

Bambang Yudhoyono said that situation is under control. The nation so panic that very under control but it has make sure the sitting well. Government has to manage the disaster. This e earthquake also make people in Thailand and India afraid. They expect that Thailand and India will flow by tsunami. And people remember tsunami in two thousand and four. Tha't the news for today. Thank you.

Appendix 5: The Holistic and Analytic Rubrics for Teachers' Rating

Holistic Rubric

Score	Level	Criteria
5	Advanced speaking	Speech is well organized in a story; information is plausible and precise and is presented logically and with appropriate transitions. Vocabulary is fully including idioms, colloquialisms, and pertinent cultural references. Good fluency and accurate pronunciation of individual sounds Most sentences have embedded more than 12 words.
4	Accomplished speaking	Speech is generally organized in a story; information is somewhat plausible and precise and is presented logically. Vocabulary is varied including idiomatic expressions and has high degree precision. Occasional non-native pronunciation errors, but the speaker is always intelligible. Each sentence has embedded clauses or phrases and contains at least 12 words.
3	Competent speaking	Speech is somewhat organized story, information maybe imprecise or implausible. Vocabulary in general is varied, including some use of idiomatic expressions. Some problems with speech rate and intonation but these do not cause serious problems with intelligibility. Each sentence has embedded clauses or phrases and contains at least 8 words.
2	Developing speaking	Speech may be insufficient and is poorly organized with basic ideas, information maybe imprecise or implausible. Numerous vocabulary words are repeated rather than using a variety of words. Numerous phonemic errors and foreign stress that cause the speaker to be occasionally unintelligible. A few sentences have embedded clauses or phrases and contain at least 5 words.
1	Beginning speaking	Limited ability to respond to the story; information is irrelevant or inaccurate Very few vocabulary words are used; single words are used rather than complete thoughts.

	Very significant phonemic errors and foreign stress that causes the speaker to be unintelligible Each sentence has no embedded clauses or phrases and contains less than 5 words.

Analytic Rubric

Category	Score	Description
Pronunciation	1	Very significant phonemic errors and foreign stress and intonation patterns so the speaker is unintelligible.
	2	Frequent phonemic errors and foreign stress and intonation patterns so speaker is a fairly intelligible.
	3	Some consistent phonemic errors and foreign stress and intonation patterns, but the speaker is intelligible.
	4	Occasional pronunciation errors, but the speaker is always intelligible.
	5	Few non-native pronunciation errors with non-native accent
Discourse	1	Limited ability to respond to the story; information may be irrelevant or inaccurate.
	2	Speech may be insufficient and is poorly organized with basic ideas; information is imprecise or inaccurate.
	3	Speech is somewhat insufficient and is poorly organized, information maybe imprecise or inaccurate.
	4	Speech is generally organized in a story; information is plausible and precise and is presented logically.
	5	Speech is well organized in a story; information is plausible and precise and is presented logically and with appropriate transitions.
Vocabulary	1	Very few vocabulary words are used; single words are used rather than complete thoughts.
	2	Numerous vocabulary words are repeated rather than using a variety of words.
	3	Vocabulary in general is varied, including some use of idiomatic expressions.
	4	Vocabulary is varied, including idiomatic expressions.
	5	Vocabulary is fully including idioms, colloquialisms, and pertinent cultural references.
Grammar	1	Virtually no grammatical or syntactical control except in simple stock phrases.
	2	Some control of basic grammatical construction but with major and/or repeated errors that interfere with intelligibility.

	3	Generally good control in all construction with grammatical errors that do not interfere with overall intelligibility.
	4	Sporadic minor grammatical errors that could be made inadvertently by native speakers.
	5	Few grammatical errors that could be made inadvertently by native speakers.
Sentence Complexity	1	Each sentence has no embedded clauses or phrases and contains less than 5 words.
	2	A few sentences have embedded clauses or phrases and contain at least 5 words.
	3	Each sentence has embedded clauses or phrases and contains at least 8 words.
	4	Each sentence has embedded clauses or phrases and contains at least 12 words.
	5	Most sentences have embedded more than 12 words.

Appendix 6: Full Descriptive Statistics for Speech Fluency Measures and Questionnaire

Descriptive Statistics for Speech Rate, Articulation Rate, Phonation Time Ratio, and Mean Length of Run									
		Speech Rate		Articulation Rate		Phonation/Time Ratio		Mean Length of Run	
Explicit Group		1	2	1	2	1	2	1	2
Mean		2.47	2.61	3.55	3.51	0.70	0.75	6.19	6.75
SEM		0.09	0.08	0.12	0.08	0.02	0.01	0.36	0.46
95% CI	Lower	2.29	2.45	3.30	3.34	0.66	0.72	5.46	5.80
	Upper	2.65	2.77	3.80	3.68	0.74	0.78	6.93	7.71
5% TM		2.47	2.61	3.53	3.50	0.71	0.75	6.05	6.48
Median		2.47	2.58	3.45	3.46	0.72	0.76	6.30	6.65
Variance		0.22	0.17	0.42	0.20	0.01	0.01	3.74	6.02
SD		0.46	0.41	0.65	0.44	0.10	0.07	1.93	2.45
Minimum		1.39	1.86	2.35	2.78	0.42	0.61	2.54	3.87
Maximum		3.60	3.37	5.34	4.58	0.89	0.92	13.43	16.36
Range		2.20	1.51	2.99	1.80	0.47	0.31	10.89	12.49
IQ Range		0.55	0.62	0.97	0.59	0.12	0.10	1.42	2.44
Skewness		0.04	0.04	0.73	0.52	-0.74	0.19	1.72	2.30
SE		0.43	0.44	0.43	0.44	0.43	0.44	0.43	0.44
Kurtosis		0.50	-0.77	0.71	0.26	1.09	0.04	6.45	7.99
SE		0.85	0.86	0.85	0.86	0.85	0.86	0.85	0.86
Implicit Group									
Mean		2.83	2.84	3.86	4.02	0.72	0.71	7.22	6.42
SEM		0.14	0.13	0.14	0.11	0.02	0.02	0.78	0.36
95% CI	Lower	2.54	2.58	3.58	3.78	0.69	0.68	5.62	5.67
	Upper	3.12	3.10	4.14	4.25	0.76	0.75	8.83	7.16
5% TM		2.81	2.81	3.84	4.01	0.73	0.71	6.76	6.33
Median		2.71	2.80	3.76	4.09	0.73	0.70	6.40	6.15
Variance		0.49	0.39	0.46	0.32	0.01	0.01	15.13	3.25
SD		0.70	0.63	0.68	0.56	0.08	0.08	3.89	1.80
Minimum		1.44	1.74	2.87	2.99	0.47	0.58	3.22	3.29
Maximum		4.54	4.54	5.17	5.14	0.85	0.88	21.33	10.93
Range		3.10	2.80	2.31	2.16	0.38	0.30	18.11	7.64
IQ Range		0.81	0.60	1.10	0.87	0.11	0.12	3.85	1.74
Skewness		0.38	0.74	0.49	-0.10	-1.01	0.55	2.09	1.10
SE		0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis		0.52	1.28	-0.67	-0.53	2.08	-0.29	6.37	1.53
SE		0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Note: 1 = Pre-test and 2 = Post-test									

Descriptive Statistics for Number of Pauses, Number of Filled Pauses, Mean Duration of Pauses, Number of Repetitions, and Number of Repairs											
		Number of Pauses		Number of Filled Pauses		Mean Duration of Pauses		Number of Repetition		Number of Repair	
Explicit Group		1	2	1	2	1	2	1	2	1	2
Mean		0.17	0.15	0.02	0.01	0.77	0.67	0.01	0.01	0.01	0.01
SEM		0.01	0.01	0.01	0.00	0.05	0.02	0.00	0.00	0.00	0.00
95% CI	Lower	0.14	0.14	0.01	0.00	0.67	0.62	0.00	0.00	0.01	0.00
	Upper	0.19	0.17	0.03	0.01	0.87	0.72	0.02	0.01	0.01	0.01
5% TM		0.16	0.15	0.02	0.01	0.76	0.67	0.01	0.01	0.01	0.01
Median		0.15	0.14	0.01	0.00	0.74	0.67	0.00	0.01	0.01	0.01
Variance		0.00	0.00	0.00	0.00	0.07	0.01	0.00	0.00	0.00	0.00
SD		0.06	0.05	0.03	0.01	0.26	0.12	0.02	0.01	0.01	0.01
Minimum		0.06	0.06	0.00	0.00	0.25	0.43	0.00	0.00	0.00	0.00
Maximum		0.38	0.25	0.11	0.03	1.64	0.92	0.09	0.03	0.03	0.03
Range		0.31	0.19	0.11	0.03	1.39	0.49	0.09	0.03	0.03	0.03
IQ Range		0.04	0.06	0.03	0.01	0.32	0.20	0.01	0.01	0.02	0.01
Skewness		1.85	0.11	1.62	1.30	1.12	-0.04	3.05	1.08	0.33	1.23
SE		0.43	0.44	0.43	0.44	0.43	0.44	0.43	0.44	0.43	0.44
Kurtosis		5.87	-0.32	2.29	0.51	3.31	-0.53	10.24	0.33	-0.33	0.91
SE		0.85	0.86	0.85	0.86	0.85	0.86	0.85	0.86	0.85	0.86
Implicit Group											
Mean		0.16	0.16	0.02	0.01	0.67	0.71	0.02	0.01	0.01	0.01
SEM		0.02	0.01	0.00	0.00	0.04	0.05	0.00	0.00	0.00	0.00
95% CI	Lower	0.13	0.14	0.01	0.00	0.60	0.61	0.01	0.01	0.00	0.00
	Upper	0.19	0.18	0.03	0.02	0.75	0.80	0.02	0.02	0.01	0.01
5% TM		0.16	0.16	0.02	0.01	0.66	0.70	0.01	0.01	0.01	0.01
Median		0.14	0.15	0.02	0.01	0.66	0.70	0.01	0.01	0.01	0.00
Variance		0.01	0.00	0.00	0.00	0.03	0.06	0.00	0.00	0.00	0.00
SD		0.08	0.05	0.02	0.02	0.18	0.24	0.02	0.01	0.01	0.01
Minimum		0.03	0.08	0.00	0.00	0.36	0.22	0.00	0.00	0.00	0.00
Maximum		0.30	0.30	0.09	0.09	1.24	1.37	0.06	0.04	0.04	0.06
Range		0.27	0.21	0.09	0.09	0.88	1.15	0.06	0.04	0.04	0.06
IQ Range		0.09	0.05	0.03	0.02	0.13	0.27	0.02	0.02	0.01	0.01
Skewness		0.58	0.96	1.36	3.70	1.61	0.70	1.39	1.67	1.43	2.77
SE		0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis		-0.48	2.45	1.72	16.27	4.26	1.76	1.56	2.64	2.22	8.81
SE		0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

Note:

1. 1 = Pre-test and 2 = Post-test
2. Number of Repetition and Repair are resulted from dividing the actual numbers and the number of Syllable
3. Number of Repair is the accumulation of number of false start, reformulation, and replacement

Descriptive Statistics of Questionnaire Responses (questions 1 – 13)													
Explicit	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**
Mean	2.07	2.79	2.00	1.69	1.41	3.10	2.48	1.45	1.83	2.52	2.62	1.79	1.38
SEM	0.19	0.15	0.19	0.15	0.12	0.19	0.21	0.12	0.16	0.20	0.25	0.15	0.16
95% CI Lower	1.69	2.48	1.62	1.38	1.18	2.72	2.06	1.21	1.51	2.10	2.12	1.48	1.05
95% CI Upper	2.45	3.10	2.38	2.00	1.65	3.49	2.91	1.69	2.15	2.93	3.12	2.10	1.71
5% TM	2.02	2.83	1.94	1.66	1.35	3.10	2.48	1.39	1.77	2.48	2.58	1.77	1.31
Median	2.00	3.00	2.00	1.00	1.00	3.00	3.00	1.00	2.00	3.00	2.00	2.00	1.00
Variance	1.00	0.67	1.00	0.65	0.39	1.02	1.26	0.40	0.72	1.19	1.74	0.67	0.74
Std. Deviation	1.00	0.82	1.00	0.81	0.63	1.01	1.12	0.63	0.85	1.09	1.32	0.82	0.86
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	0.00
Maximum	4.00	4.00	4.00	3.00	3.00	5.00	5.00	3.00	4.00	5.00	5.00	3.00	4.00
Range	3.00	3.00	3.00	2.00	2.00	4.00	5.00	2.00	3.00	4.00	4.00	2.00	4.00
IQ Range	2.00	0.50	2.00	1.00	1.00	2.00	1.00	1.00	1.00	1.00	2.50	1.50	0.50
Skewness	0.32	-0.85	0.46	0.65	1.28	0.00	-0.12	1.12	0.73	0.22	0.27	0.41	1.66
SE	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Kurtosis	-1.16	0.70	-1.05	-1.14	0.69	-0.69	-0.01	0.31	-0.15	-0.52	-1.27	-1.38	2.44
SE	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Implicit													
Mean	2.20	2.72	1.48	1.64	1.68	2.32	2.72	1.32	1.52	2.44	2.28	1.68	1.64
SEM	0.19	0.14	0.14	0.17	0.16	0.21	0.19	0.14	0.10	0.20	0.20	0.15	0.14
95% CI Lower	1.80	2.44	1.19	1.28	1.35	1.88	2.33	1.04	1.31	2.03	1.86	1.37	1.35
95% CI Upper	2.60	3.00	1.77	2.00	2.01	2.76	3.11	1.60	1.73	2.85	2.70	1.99	1.93
5% TM	2.17	2.76	1.42	1.56	1.64	2.26	2.74	1.24	1.52	2.43	2.26	1.64	1.60
Median	2.00	3.00	1.00	1.00	1.00	2.00	3.00	1.00	2.00	2.00	2.00	2.00	2.00
Variance	0.92	0.46	0.51	0.74	0.64	1.14	0.88	0.48	0.26	1.01	1.04	0.56	0.49
Std. Deviation	0.96	0.68	0.71	0.86	0.80	1.07	0.94	0.69	0.51	1.00	1.02	0.75	0.70
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	4.00	4.00	3.00	4.00	3.00	5.00	4.00	3.00	2.00	4.00	4.00	3.00	3.00
Range	3.00	3.00	2.00	3.00	2.00	4.00	3.00	2.00	1.00	3.00	3.00	2.00	2.00
IQ Range	2.00	0.50	1.00	1.00	1.00	2.00	1.00	0.00	1.00	1.00	2.00	1.00	1.00
Skewness	0.19	-1.33	1.19	1.23	0.67	0.40	-0.37	1.95	-0.09	0.04	0.15	0.62	0.64
SE	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis	-0.94	1.99	0.15	0.86	-1.08	-0.02	-0.53	2.36	-2.17	-0.97	-1.09	-0.89	-0.64
SE	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

Descriptive Statistics of Questionnaire Responses (questions 14 – 26)													
Explicit	**Q14**	**Q15**	**Q16**	**Q17**	**Q18**	**Q19**	**Q20**	**Q21**	**Q22**	**Q23**	**Q24**	**Q25**	**Q26**
Mean	2.00	1.41	2.69	2.79	1.90	1.83	2.41	2.45	2.17	2.07	2.52	1.97	1.41
SEM	0.16	0.11	0.11	0.18	0.16	0.12	0.18	0.20	0.17	0.18	0.23	0.25	0.12
95% CI Lower	1.68	1.20	2.46	2.42	1.57	1.58	2.05	2.03	1.82	1.70	2.06	1.46	1.18
95% CI Upper	2.32	1.63	2.92	3.16	2.22	2.08	2.77	2.86	2.53	2.43	2.98	2.47	1.65
5% TM	1.96	1.37	2.66	2.83	1.85	1.81	2.40	2.44	2.14	2.02	2.48	1.85	1.35
Median	2.00	1.00	3.00	3.00	2.00	2.00	3.00	2.00	2.00	2.00	3.00	1.00	1.00
Variance	0.71	0.32	0.36	0.96	0.74	0.43	0.89	1.18	0.86	0.92	1.47	1.75	0.39
Std. Deviation	0.85	0.57	0.60	0.98	0.86	0.66	0.95	1.09	0.93	0.96	1.21	1.32	0.63
Minimum	1.00	1.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	4.00	3.00	4.00	4.00	4.00	3.00	4.00	4.00	4.00	4.00	5.00	5.00	3.00
Range	3.00	2.00	2.00	3.00	3.00	2.00	3.00	3.00	3.00	3.00	4.00	4.00	2.00
IQ Range	2.00	1.00	1.00	0.50	1.50	1.00	1.00	1.50	2.00	2.00	2.50	2.00	1.00
Skewness	0.38	0.99	0.24	-0.79	0.57	0.19	-0.14	0.05	0.21	0.37	0.15	1.16	1.28
SE	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Kurtosis	-0.57	0.07	-0.51	-0.19	-0.47	-0.57	-0.88	-1.25	-0.87	-0.92	-1.10	0.10	0.69
SE	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Implicit													
Mean	2.04	1.48	2.88	2.40	1.84	2.08	2.64	2.52	2.60	1.92	3.12	1.68	1.28
SEM	0.17	0.15	0.16	0.19	0.18	0.16	0.14	0.17	0.20	0.16	0.19	0.17	0.09
95% CI Lower	1.69	1.16	2.56	2.00	1.47	1.74	2.35	2.16	2.19	1.58	2.72	1.33	1.09
95% CI Upper	2.39	1.80	3.20	2.80	2.21	2.42	2.93	2.88	3.01	2.26	3.52	2.03	1.47
5% TM	2.04	1.38	2.87	2.39	1.82	2.09	2.64	2.52	2.61	1.87	3.12	1.64	1.26
Median	2.00	1.00	3.00	2.00	2.00	2.00	3.00	3.00	3.00	2.00	3.00	1.00	1.00
Variance	0.71	0.59	0.61	0.92	0.81	0.66	0.49	0.76	1.00	0.66	0.94	0.73	0.21
Std. Deviation	0.84	0.77	0.78	0.96	0.90	0.81	0.70	0.87	1.00	0.81	0.97	0.85	0.46
Minimum	1.00	1.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	3.00	4.00	4.00	4.00	3.00	3.00	4.00	4.00	4.00	4.00	5.00	3.00	2.00
Range	2.00	3.00	2.00	3.00	2.00	2.00	3.00	3.00	3.00	3.00	4.00	2.00	1.00
IQ Range	2.00	1.00	1.50	1.00	2.00	2.00	1.00	1.00	1.00	1.00	1.50	1.50	1.00
Skewness	-0.08	1.86	0.22	0.00	0.34	-0.15	-0.15	-0.07	-0.16	0.66	0.04	0.70	1.04
SE	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis	-1.59	3.73	-1.28	-0.85	-1.73	-1.45	0.11	-0.49	-0.92	0.26	-0.01	-1.26	-1.00
SE	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

Descriptive Statistics of Questionnaire Responses (questions 27 – 39)													
Explicit	**Q27**	**Q28**	**Q29**	**Q30**	**Q31**	**Q32**	**Q33**	**Q34**	**Q35**	**Q36**	**Q37**	**Q38**	**Q39**
Mean	1.21	2.17	1.59	1.52	1.86	2.34	1.45	1.83	2.48	1.90	1.66	2.00	1.97
SEM	0.08	0.19	0.19	0.15	0.12	0.14	0.14	0.16	0.26	0.17	0.13	0.18	0.14
95% CI Lower	1.05	1.78	1.20	1.22	1.62	2.05	1.17	1.51	1.96	1.54	1.38	1.63	1.67
95% CI Upper	1.36	2.57	1.97	1.82	2.11	2.64	1.73	2.15	3.01	2.25	1.93	2.37	2.26
5% TM	1.17	2.10	1.48	1.43	1.85	2.38	1.43	1.77	2.43	1.83	1.62	1.96	1.96
Median	1.00	2.00	1.00	1.00	2.00	3.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00
Variance	0.17	1.08	1.04	0.62	0.41	0.59	0.54	0.72	1.90	0.88	0.52	0.93	0.61
Std. Deviation	0.41	1.04	1.02	0.78	0.64	0.77	0.74	0.85	1.38	0.94	0.72	0.96	0.78
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	2.00	5.00	4.00	4.00	3.00	3.00	3.00	4.00	5.00	4.00	3.00	4.00	3.00
Range	1.00	4.00	3.00	3.00	2.00	2.00	3.00	3.00	4.00	3.00	2.00	3.00	2.00
IQ Range	0.00	2.00	1.50	1.00	1.00	1.00	1.00	1.00	3.00	1.50	1.00	2.00	2.00
Skewness	1.53	0.66	1.38	1.61	0.12	-0.70	0.77	0.73	0.52	0.77	0.64	0.26	0.06
SE	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Kurtosis	0.35	0.33	0.37	2.41	-0.40	-0.91	0.15	-0.15	-1.01	-0.26	-0.76	-1.40	-1.30
SE	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Implicit													
Mean	1.36	1.72	1.64	1.32	2.36	2.64	1.32	2.16	2.32	1.72	1.96	2.24	2.00
SEM	0.13	0.17	0.15	0.10	0.23	0.15	0.11	0.16	0.17	0.16	0.18	0.14	0.15
95% CI Lower	1.10	1.37	1.33	1.12	1.89	2.33	1.09	1.83	1.97	1.39	1.59	1.94	1.68
95% CI Upper	1.62	2.07	1.95	1.52	2.83	2.95	1.55	2.49	2.67	2.05	2.33	2.54	2.32
5% TM	1.29	1.69	1.60	1.30	2.29	2.66	1.26	2.18	2.30	1.69	1.96	2.27	2.00
Median	1.00	1.00	1.00	1.00	2.00	3.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00
Variance	0.41	0.71	0.57	0.23	1.32	0.57	0.31	0.64	0.73	0.63	0.79	0.52	0.58
Std. Deviation	0.64	0.84	0.76	0.48	1.15	0.76	0.56	0.80	0.85	0.79	0.89	0.72	0.76
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	3.00	3.00	3.00	2.00	5.00	4.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00
Range	2.00	2.00	2.00	1.00	4.00	3.00	2.00	2.00	3.00	2.00	2.00	2.00	2.00
IQ Range	1.00	1.50	1.00	1.00	2.00	1.00	1.00	1.50	1.00	1.00	2.00	1.00	2.00
Skewness	1.62	0.60	0.73	0.82	0.64	-0.52	1.58	-0.31	0.18	0.56	0.08	-0.41	0.00
SE	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis	1.64	-1.33	-0.81	-1.45	0.34	0.28	1.84	-1.34	-0.38	-1.14	-1.78	-0.91	-1.21
SE	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

Descriptive Statistics of Questionnaire Responses (questions 40 – 52)													
Explicit	**Q40**	**Q41**	**Q42**	**Q43**	**Q44**	**Q45**	**Q46**	**Q47**	**Q48**	**Q49**	**Q50**	**Q51**	**Q52**
Mean	1.45	2.52	2.48	2.41	2.31	2.52	1.86	1.38	1.76	1.38	2.07	1.79	1.28
SEM	0.13	0.17	0.21	0.19	0.19	0.21	0.16	0.10	0.17	0.12	0.16	0.18	0.08
95% CI Lower	1.19	2.17	2.04	2.03	1.93	2.08	1.53	1.17	1.41	1.14	1.73	1.42	1.10
95% CI Upper	1.71	2.86	2.92	2.80	2.69	2.96	2.19	1.59	2.11	1.62	2.41	2.16	1.45
5% TM	1.39	2.52	2.44	2.37	2.29	2.48	1.81	1.33	1.73	1.31	2.04	1.71	1.25
Median	1.00	3.00	3.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	2.00	1.00	1.00
Variance	0.47	0.83	1.33	1.04	1.01	1.33	0.77	0.32	0.83	0.39	0.78	0.96	0.21
Std. Deviation	0.69	0.91	1.15	1.02	1.00	1.15	0.88	0.56	0.91	0.62	0.88	0.98	0.45
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	3.00	4.00	5.00	5.00	4.00	5.00	4.00	3.00	3.00	3.00	4.00	4.00	2.00
Range	2.00	3.00	4.00	4.00	3.00	4.00	3.00	2.00	2.00	2.00	3.00	3.00	1.00
IQ Range	1.00	1.00	2.00	1.00	2.00	1.50	1.50	1.00	2.00	1.00	2.00	1.50	1.00
Skewness	1.27	-0.05	0.12	0.80	0.00	0.33	0.63	1.16	0.52	1.45	0.19	0.94	1.06
SE	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Kurtosis	0.41	-0.66	-0.82	0.23	-1.14	-0.85	-0.52	0.49	-1.65	1.16	-1.00	-0.25	-0.95
SE	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Implicit													
Mean	1.36	2.44	2.72	2.32	2.72	2.60	1.80	1.16	1.52	1.16	1.80	1.32	1.28
SEM	0.11	0.14	0.19	0.19	0.19	0.22	0.15	0.09	0.13	0.07	0.17	0.15	0.11
95% CI Lower	1.13	2.15	2.33	1.93	2.33	2.14	1.48	0.96	1.25	1.01	1.44	1.01	1.06
95% CI Upper	1.59	2.73	3.11	2.71	3.11	3.06	2.12	1.36	1.79	1.31	2.16	1.63	1.50
5% TM	1.30	2.49	2.70	2.30	2.70	2.56	1.78	1.08	1.47	1.12	1.73	1.20	1.21
Median	1.00	3.00	3.00	2.00	3.00	2.00	2.00	1.00	1.00	1.00	2.00	1.00	1.00
Variance	0.32	0.51	0.88	0.89	0.88	1.25	0.58	0.22	0.43	0.14	0.75	0.56	0.29
Std. Deviation	0.57	0.71	0.94	0.95	0.94	1.12	0.76	0.47	0.65	0.37	0.87	0.75	0.54
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	3.00	3.00	5.00	4.00	5.00	5.00	3.00	3.00	3.00	2.00	4.00	4.00	3.00
Range	2.00	2.00	4.00	3.00	4.00	4.00	2.00	2.00	2.00	1.00	3.00	3.00	2.00
IQ Range	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.50
Skewness	1.34	-0.90	0.29	0.25	-0.04	0.70	0.37	3.14	0.90	1.98	0.84	2.62	1.86
SE	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis	1.04	-0.38	0.40	-0.69	0.77	-0.02	-1.14	9.97	-0.15	2.06	0.03	6.89	2.94
SE 0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	

Descriptive Statistics of Questionnaire Responses (questions 53 – 66/ finish)														
Explicit	**Q53**	**Q54**	**Q55**	**Q56**	**Q57**	**Q58**	**Q59**	**Q60**	**Q61**	**Q62**	**Q63**	**Q64**	**Q65**	**Q66**
Mean	2.45	2.45	2.17	2.48	1.55	2.03	2.17	2.00	2.24	1.93	1.72	1.21	2.07	3.00
SEM	0.17	0.21	0.23	0.16	0.17	0.15	0.17	0.19	0.12	0.16	0.16	0.08	0.18	0.18
95% CI Lower	2.10	2.02	1.71	2.15	1.21	1.72	1.82	1.61	2.00	1.59	1.39	1.05	1.70	2.63
95% CI Upper	2.79	2.87	2.64	2.81	1.90	2.35	2.53	2.39	2.48	2.27	2.06	1.36	2.43	3.37
5% TM	2.42	2.40	2.10	2.48	1.46	2.00	2.17	1.94	2.27	1.87	1.66	1.17	2.02	3.02
Median	3.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	2.00	3.00
Variance	0.83	1.26	1.50	0.76	0.83	0.68	0.86	1.07	0.40	0.78	0.78	0.17	0.92	0.93
Std. Deviation	0.91	1.12	1.23	0.87	0.91	0.82	0.93	1.04	0.64	0.88	0.88	0.41	0.96	0.96
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	5.00	5.00	5.00	4.00	4.00	4.00	4.00	4.00	3.00	4.00	4.00	2.00	4.00	5.00
Range	4.00	4.00	4.00	3.00	3.00	3.00	4.00	3.00	2.00	3.00	3.00	1.00	3.00	4.00
IQ Range	1.00	1.50	2.00	1.00	1.00	2.00	1.00	2.00	1.00	1.00	1.00	0.00	2.00	2.00
Skewness	0.16	0.30	0.77	0.23	1.36	0.35	0.21	0.42	-0.24	0.81	0.93	1.53	0.63	-0.26
SE	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Kurtosis	0.92	-0.61	-0.58	-0.50	0.52	-0.44	0.56	-1.27	-0.50	0.24	-0.12	0.35	-0.38	-0.17
SE	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Implicit														
Mean	2.56	2.96	1.84	2.40	1.28	1.88	1.84	1.48	2.04	1.80	1.68	1.24	1.56	3.00
SEM	0.14	0.20	0.17	0.15	0.15	0.16	0.17	0.12	0.15	0.15	0.16	0.10	0.14	0.17
95% CI Lower	2.27	2.54	1.49	2.08	0.98	1.56	1.49	1.24	1.74	1.48	1.35	1.02	1.27	2.64
95% CI Upper	2.85	3.38	2.19	2.72	1.58	2.20	2.19	1.72	2.34	2.12	2.01	1.46	1.85	3.36
5% TM	2.62	2.96	1.78	2.40	1.16	1.87	1.78	1.43	2.04	1.78	1.60	1.17	1.51	3.00
Median	3.00	3.00	2.00	2.00	1.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	1.00	3.00
Variance	0.51	1.04	0.72	0.58	0.54	0.61	0.72	0.34	0.54	0.58	0.64	0.27	0.51	0.75
Std. Deviation	0.71	1.02	0.85	0.76	0.74	0.78	0.85	0.59	0.73	0.76	0.80	0.52	0.71	0.87
Minimum	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	3.00	5.00	4.00	4.00	4.00	3.00	4.00	3.00	3.00	3.00	4.00	3.00	3.00	5.00
Range	2.00	4.00	3.00	3.00	3.00	2.00	3.00	2.00	2.00	2.00	3.00	2.00	2.00	4.00
IQ Range	1.00	1.50	1.00	1.00	0.00	1.50	1.00	1.00	1.50	1.00	1.00	0.00	1.00	1.00
Skewness	-1.36	0.09	0.77	-0.24	2.88	0.22	0.77	0.76	-0.06	0.37	1.20	2.20	0.90	0.00
SE	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
Kurtosis	0.52	0.20	0.06	-0.31	8.22	-1.28	0.06	-0.32	-1.04	-1.14	1.46	4.46	-0.38	0.58
SE	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90

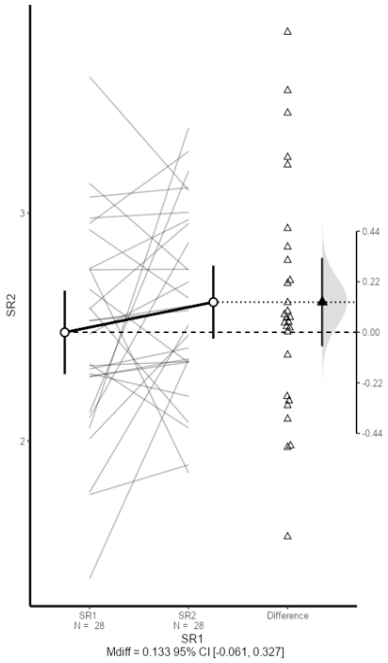
Descriptive Statistics of Raters' Judgments

Descriptive Statistics of Raters' Judgments												
Rater 1	**Proficiency**		**Pronunciation**		**Discourse**		**Vocabulary**		**Grammar**		**Complexity**	
	1	2	1	2	1	2	1	2	1	2	1	2
Mean	3.06	3.57	3.43	3.35	3.17	3.37	3.06	3.52	3.06	3.04	2.98	3.31
SEM	0.13	0.15	0.13	0.15	0.11	0.12	0.12	0.12	0.10	0.10	0.11	0.10
95% CI Lower	2.80	3.28	3.17	3.06	2.95	3.14	2.82	3.27	2.85	2.83	2.76	3.12
95% CI Upper	3.31	3.87	3.68	3.65	3.38	3.60	3.29	3.77	3.26	3.24	3.20	3.51
5% TM	3.01	3.62	3.42	3.38	3.16	3.43	3.02	3.56	3.04	3.08	2.98	3.39
Median	3.00	4.00	4.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00	3.00	3.00
Variance	0.85	1.19	0.85	1.18	0.63	0.73	0.73	0.82	0.58	0.56	0.66	0.52
Std. Deviation	0.92	1.09	0.92	1.08	0.80	0.85	0.86	0.91	0.76	0.75	0.81	0.72
Minimum	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00
Maximum	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	4.00	4.00	4.00
Range	3.00	5.00	3.00	5.00	3.00	5.00	3.00	5.00	3.00	4.00	2.00	4.00
IQ Range	2.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.25	0.25	2.00	1.00
Skewness	0.34	-0.69	-0.22	-0.29	-0.08	-1.38	0.27	-0.77	0.17	-1.17	0.03	-1.81
SE	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Kurtosis	-0.90	0.68	-0.88	0.32	-0.92	3.27	-0.80	3.12	-0.58	3.82	-1.48	7.06
SE	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
Rater 2												
Mean	0.11	3.83	4.13	4.26	3.85	4.37	2.93	3.57	4.07	4.46	3.07	3.80
SEM	0.19	0.09	0.16	0.15	0.12	0.12	0.11	0.10	0.11	0.13	0.10	0.12
95% CI Lower	2.84	3.64	3.82	3.96	3.61	4.13	2.70	3.36	3.84	4.21	2.87	3.56
95% CI Upper	3.27	4.02	4.44	4.56	4.10	4.62	3.15	3.78	4.30	4.72	3.28	4.04
5% TM	3.10	3.88	4.24	4.38	3.90	4.47	2.96	3.66	4.14	4.60	3.12	3.90
Median	3.00	4.00	4.50	5.00	4.00	5.00	3.00	4.00	4.00	5.00	3.00	4.00
Variance	0.62	0.48	1.32	1.18	0.81	0.80	0.67	0.59	0.71	0.86	0.56	0.77
Std. Deviation	0.79	0.69	1.15	1.08	0.90	0.90	0.82	0.77	0.84	0.93	0.75	0.88
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	4.00	5.00	5.00	5.00	5.00	5.00	4.00	5.00	5.00	5.00	4.00	5.00
Range	4.00	5.00	5.00	5.00	5.00	5.00	4.00	5.00	5.00	5.00	4.00	5.00
IQ Range	1.00	0.00	1.00	1.00	1.00	1.00	1.25	1.00	1.00	1.00	1.00	0.00
Skewness	-1.06	-3.29	-1.50	-1.74	-1.32	-2.45	-0.71	-2.21	-2.10	-2.70	-1.24	-3.07
SE	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Kurtosis	2.89	17.39	2.09	3.45	4.90	9.50	1.52	7.94	9.17	9.79	4.15	12.07
SE	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
Note: 1 = Pre-test, 2 = Post-test												

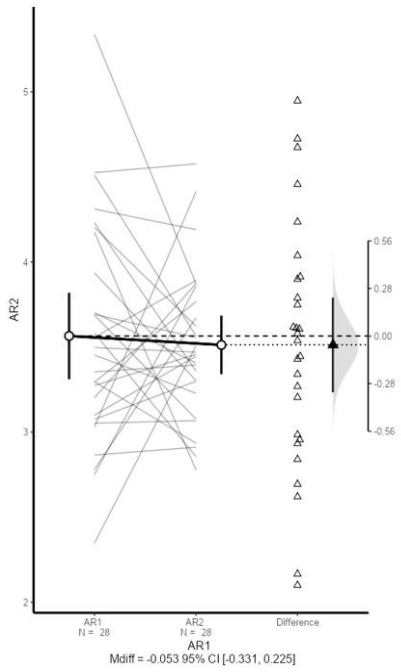
Appendix 7: Scatter Plots of All the Estimation Analysis

1. Scatter plots of the the explicit group's speech fluency measures for pre-test and post-test

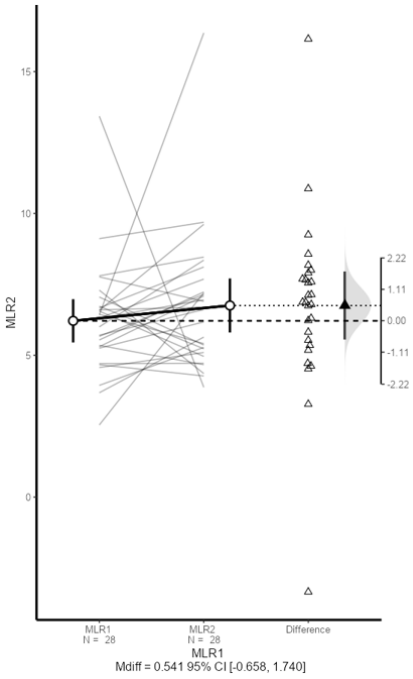
Speech rate



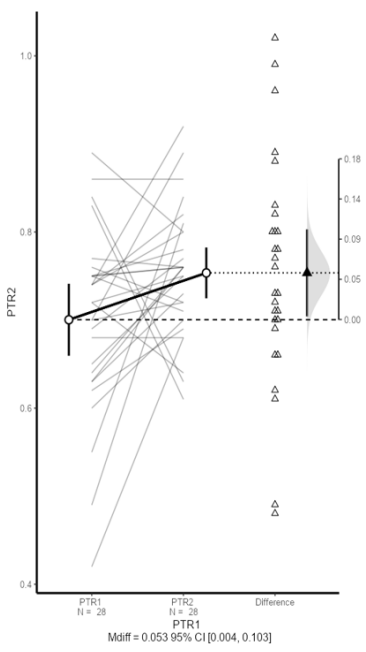
Articulation rate



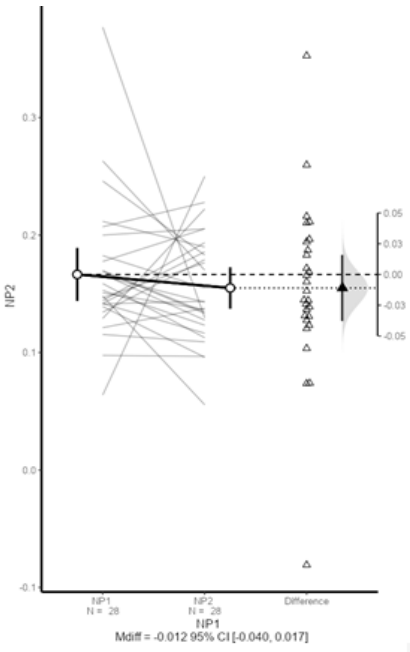
Mean length of run



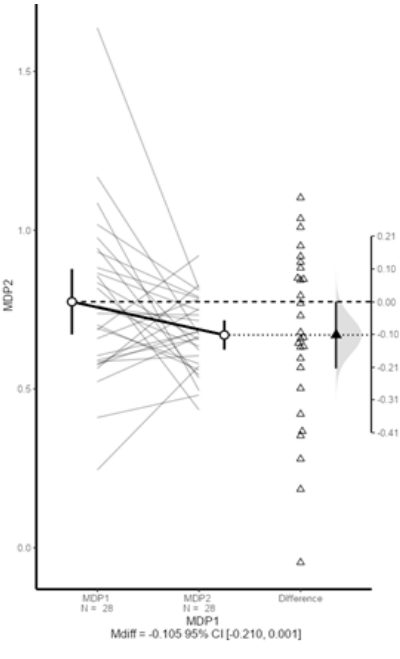
Phonation time ratio



Number of pauses

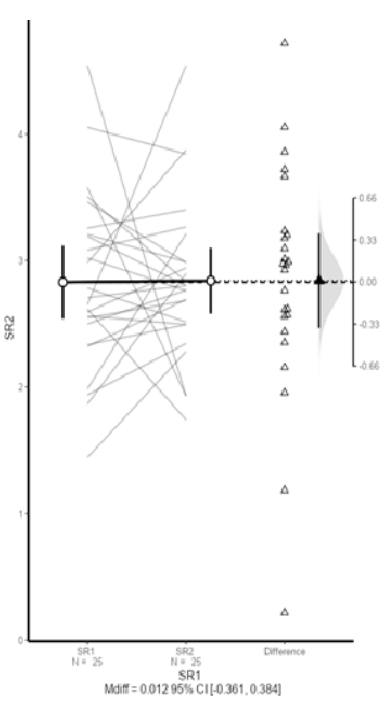


Mean duration of pauses

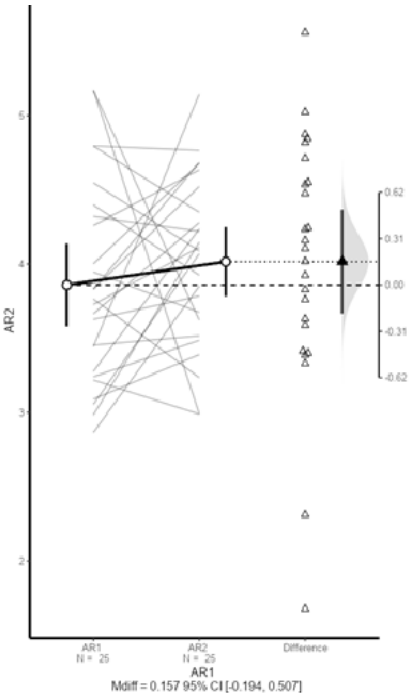


2. Scatter plots of the the implicit group's speech fluency measures for pre-test and post-test

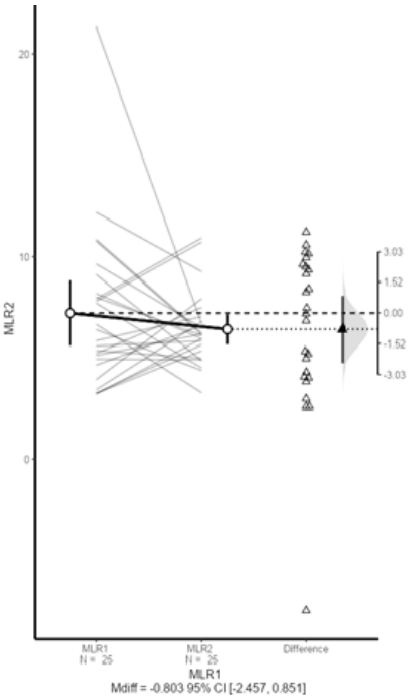
Speech rate



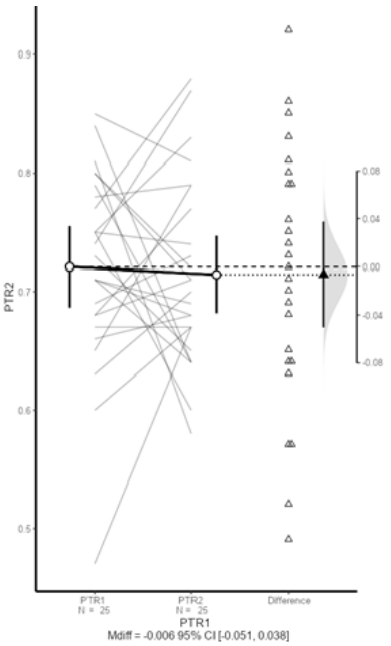
Articulation rate



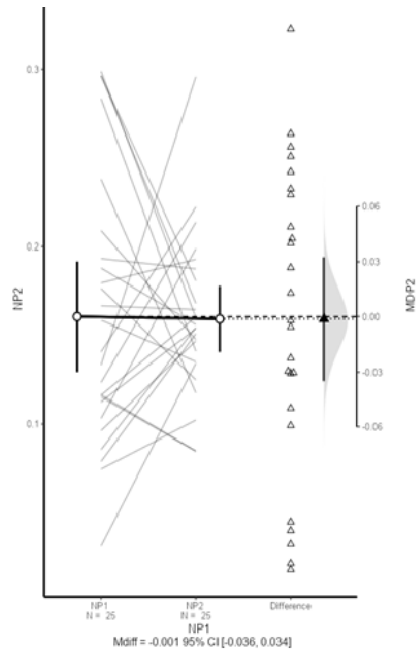
Mean length of run



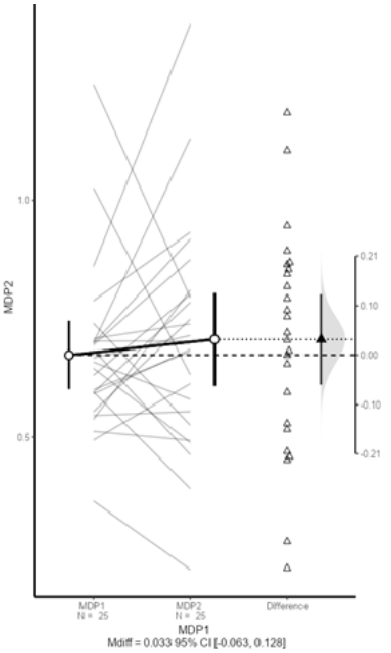
Phonation time ratio



Number of pauses

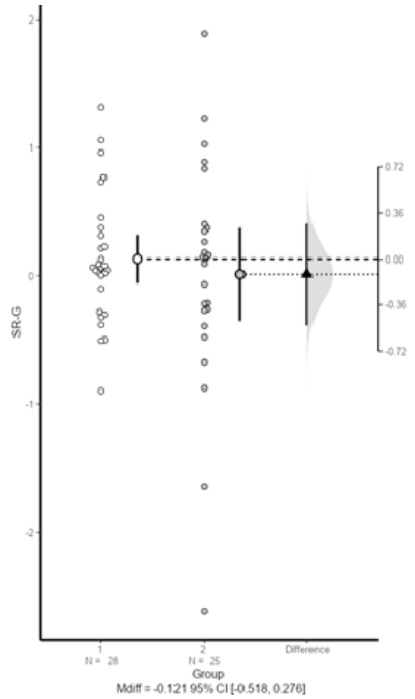


Mean duration of pauses

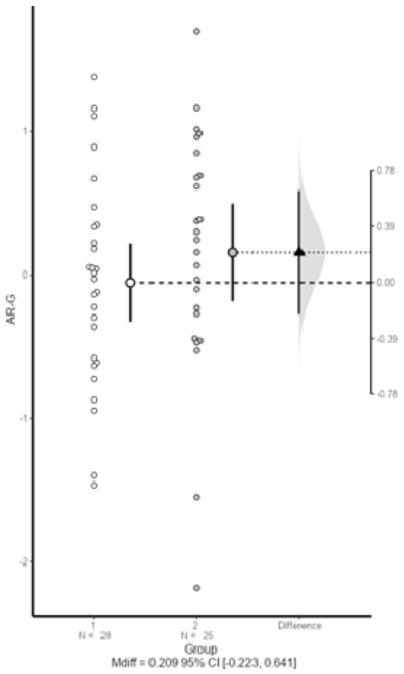


3. Scatter plots of the the group's comparison on speech fluency gains

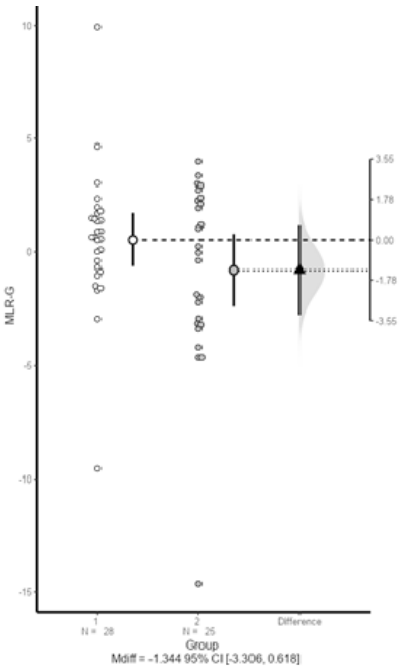
Speech rate



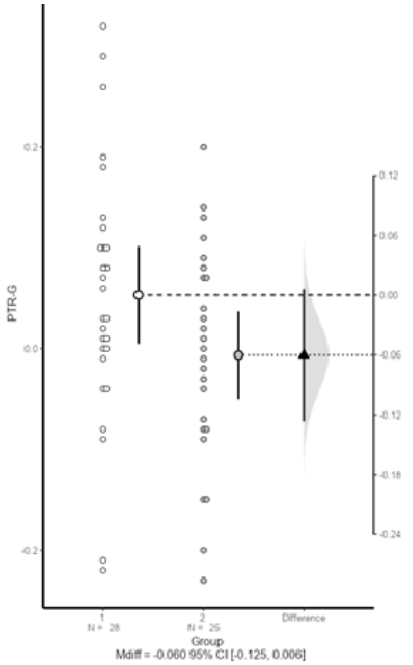
Articulation rate



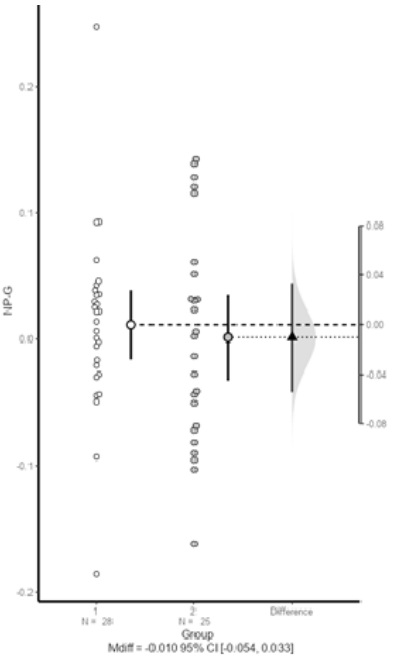
Mean length of run



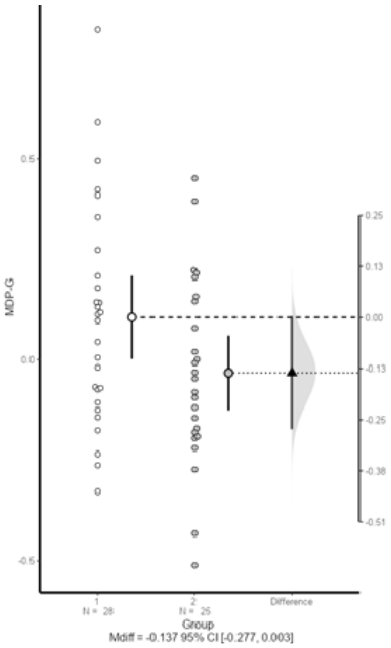
Phonation time ratio



Number of pauses

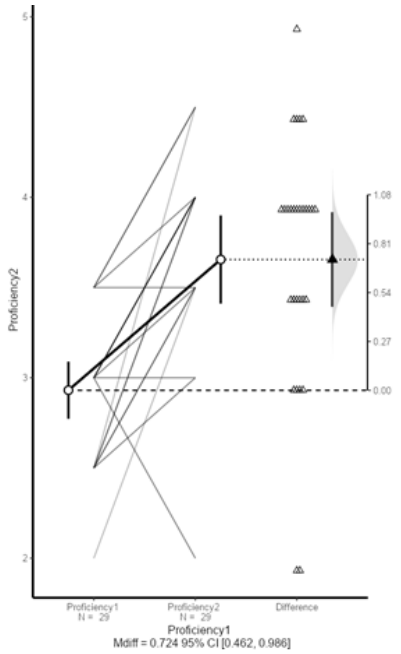


Mean duration of pauses

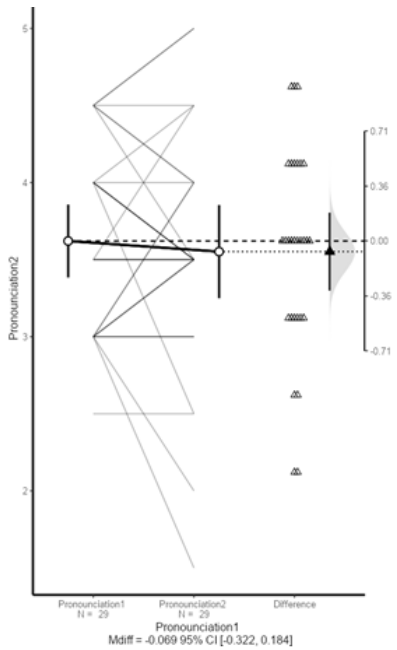


4. Scatter plots of the the explicit group's oral proficiency scales for pre-test and post-test

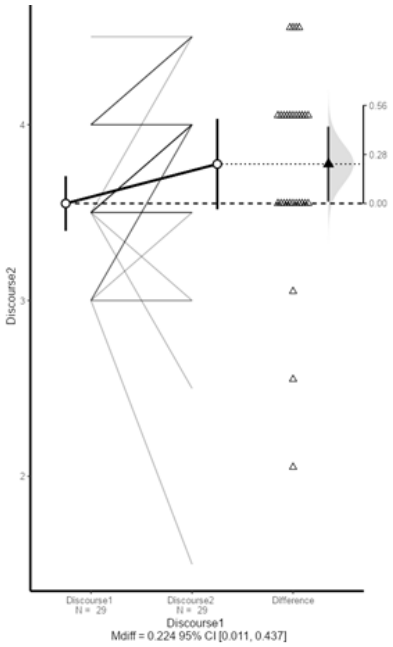
Proficiency



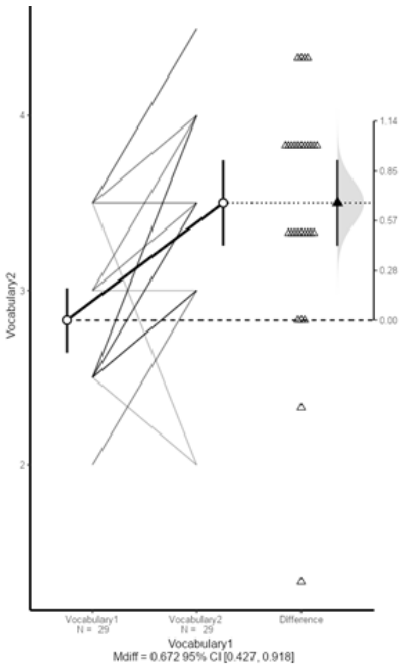
Pronunciation



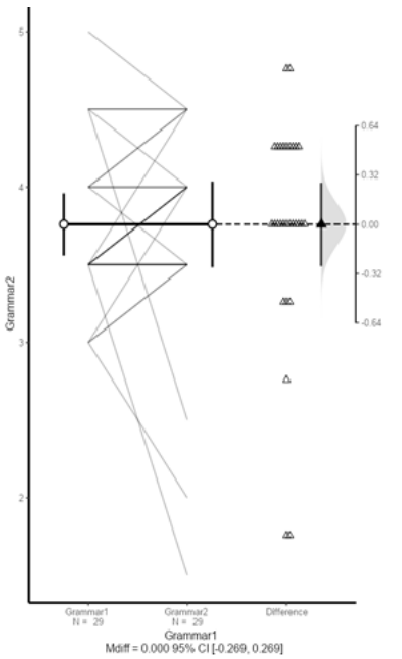
Discourse



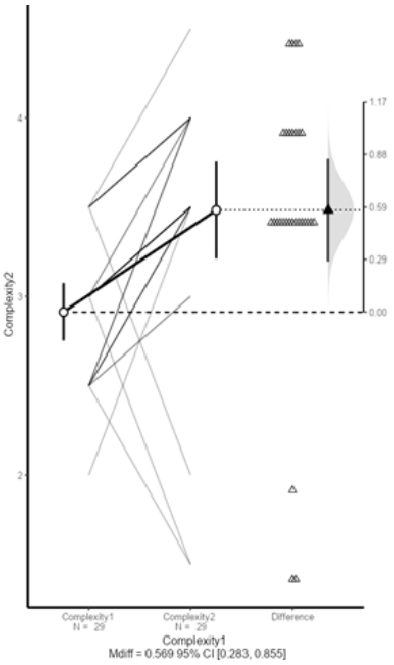
Vocabulary



Grammar

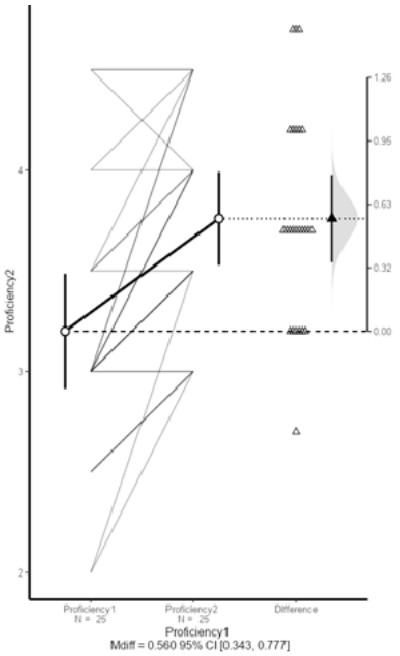


Complexity

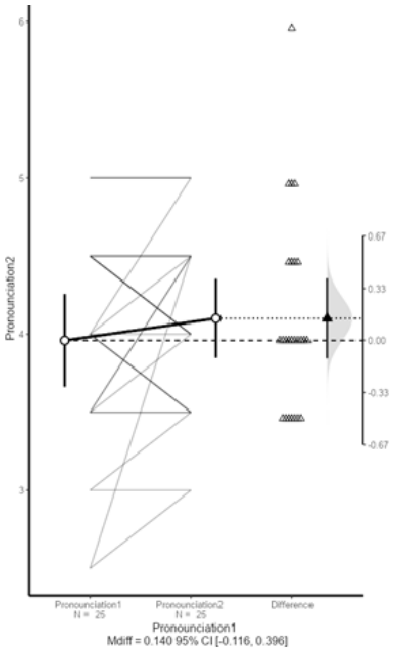


5. Scatter plots of the the implicit group's oral proficiency scales for pre-test and post-test

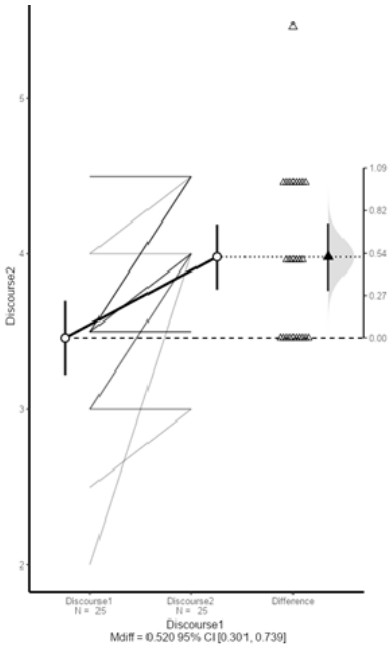
Proficiency



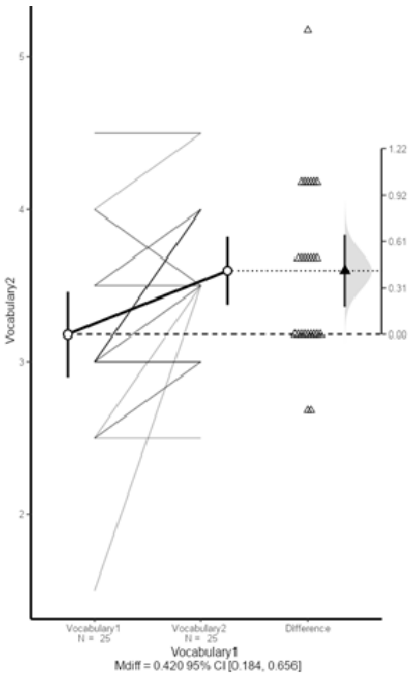
Pronunciation



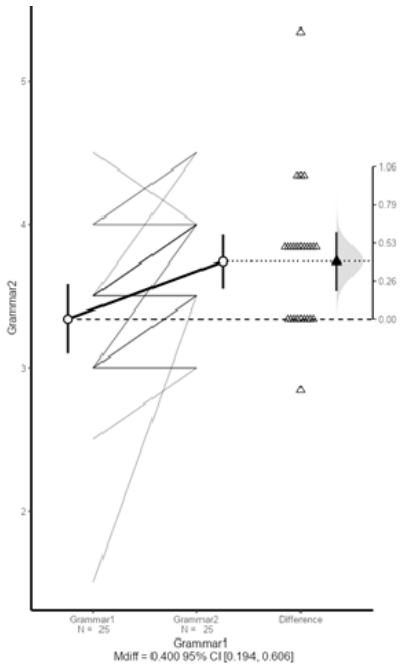
Discourse



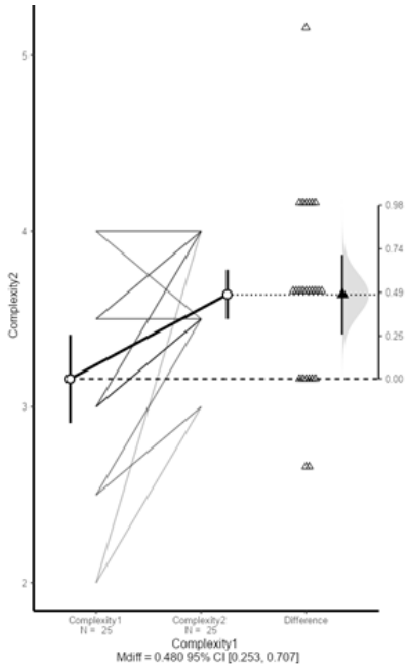
Vocabulary



Grammar

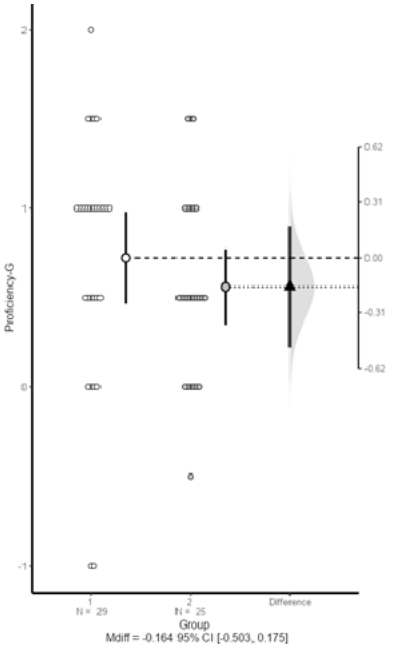


Complexity

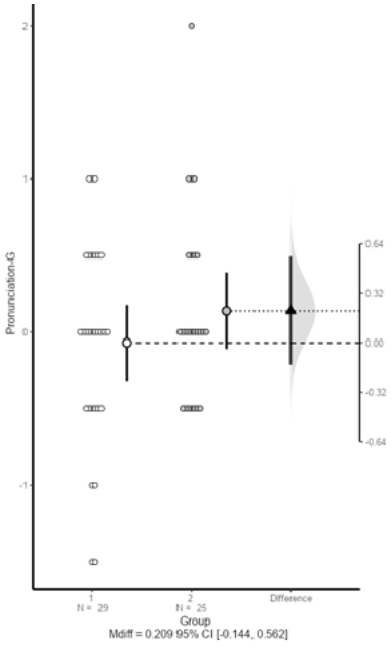


6. Scatter plots of the the group's comparison on oral proficiency gains

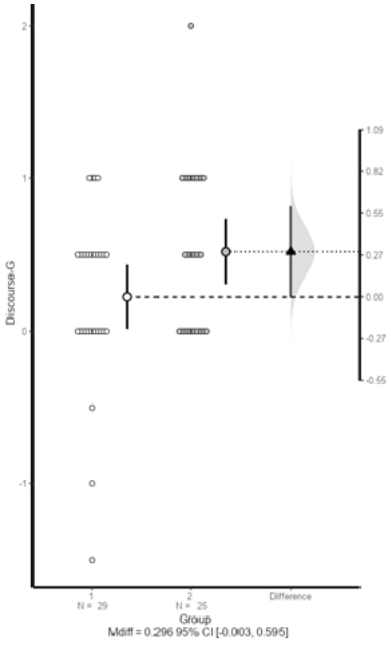
Proficiency



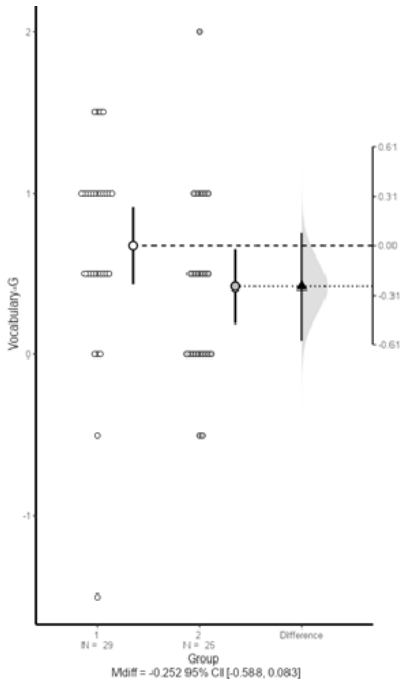
Pronunciation



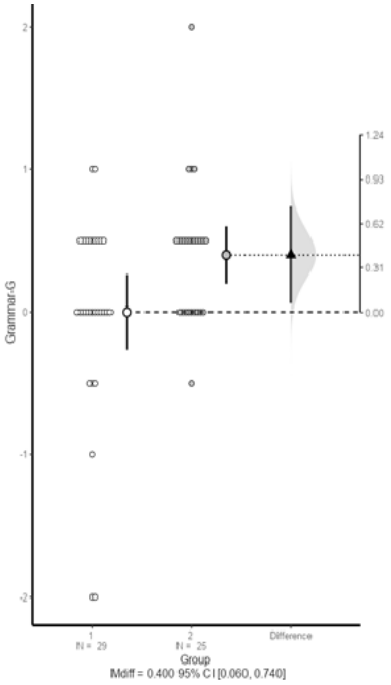
Discourse



Vocabulary



Grammar



Complexity