FLUENCY IN SIMULTANEOUS INTERPRETING OF TRAINEE INTERPRETERS: THE PERSPECTIVES OF COGNITIVE, UTTERANCE AND PERCEIVED FLUENCY

SHUXIAN SONG

PhD

The Hong Kong Polytechnic University

2020

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Fluency in Simultaneous Interpreting of Trainee Interpreters: The Perspectives of Cognitive, Utterance and Perceived Fluency

Shuxian Song

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

August, 2019

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________________ (Signed)

SHUXIAN SONG (Name of student)

Abstract

Fluency, one of the most important criteria in interpreting quality assessment and perception, has been under-explored in interpreting research, and constructs of and contributors to fluency in interpreting performance have not been fully explored. Based on the three domains of fluency proposed by Segalowitz (2010), this research offers a multidimensional exploration of fluency in the simultaneous interpreting (SI) of trainee interpreters from the perspectives of cognitive, utterance and perceived fluency. Cognitive fluency refers to the speaker's efficient mobilisation and integration of underlying cognitive processes responsible for utterance production; utterance fluency refers to the set of objectively determined oral features of utterances, e.g. temporal, hesitation and repair features; and perceived fluency refers to listeners' inferences about a speaker's cognitive fluency based on their perception of the speaker's utterance fluency.

Main issues explored in this research include: 1) the role of cognitive fluency in the fluency development of trainee interpreters' L2 (English)-L1 (Chinese) SI output under conditions of low and high cognitive load; 2) the influence of cognitive fluency, SI training and input rate on trainee interpreters' SI utterance fluency, comparing measures of speed, breakdown and repair fluency at two time slots of SI training and under conditions of low and high input rate; and 3) the relationship between objective indicators of utterance fluency in trainee interpreters' SI performance and perceived fluency, as assessed by human raters.

Cognitive fluency measures, operationalised in the current research as the efficiency (coefficient of variance) of lexical access, lexical retrieval, linguistic attention control and working memory capacity, were elicited through behavioural experiments including a semantic classification task, a word translation task, a category judgment task and a speaking span task. Measures of utterance fluency in SI performance were obtained through simulated SI tasks, which followed a 2 (training: pre/post) * 2 (input rate: low/high) factorial design. Twenty-eight trainee interpreters at the initial stage of SI training were recruited as participants. The participants interpreted two speeches simultaneously: one with a high input rate and the other with a low input rate at the beginning and end of an SI training period of one academic term. A bilingual corpus of the SI output of participants was built and indicators of SI utterance fluency were annotated systematically by using Elan 5.2 software. Raters with professional interpreting experience were invited to assess the interpreting performance.

Results of multiple linear regression analyses and repeated measures ANOVA indicated that: 1) The explored cognitive fluency measures could predict to a large extent the development of trainee interpreters' SI utterance fluency over a period of thirteen weeks; the influence of cognitive fluency measures on SI utterance fluency development was evidently stronger under conditions of high cognitive load; the efficiency of cognitive processes involved in the target language production stage had a more significant influence on SI utterance fluency development than that involved in the source language comprehension. 2) Measures of cognitive fluency in the target language generally had a significant influence on one or more indicators of utterance fluency in trainee interpreters' SI output and the role of cognitive fluency was generally independent; the main effects of SI training were significant for speech rate, articulation rate and mean duration of silent pauses, and the SI output after training was generally

more fluent than that before training; the main effects of input rate were strongly significant for speed fluency indicators and were overall significant for breakdown fluency indicators, and the fluency of trainee interpreters' SI output was generally enhanced under conditions of high input rate. 3) Objectively measured utterance fluency indicators could account for most of the variance in perceived fluency of trainee interpreters' SI performance; speech rate made a major positive contribution in predicting fluency ratings; the mean number of REPAIRs (repairs, repetitions and false starts) and the mean number of filled pauses were the second most significant predictor for perceived fluency in pre-training and post-training tasks, respectively, and were negatively correlated with fluency ratings.

The research offers an interdisciplinary exploration of fluency in SI. Exploration of the relationships between three domains of fluency in interpreting offers insights into constructs of fluency in interpreting and enriches the existing knowledge of fluency from multiple perspectives, contributing to the understanding of the information processing mechanism of interpreting. The inclusion of lexical retrieval and working memory capacity in the investigation into cognitive fluency has implications for the theoretical framework of cognitive fluency in interpreting. The focus on the efficiency of cognitive fluency and the development of utterance fluency in interpreting provides methodological references for future relevant studies. The findings also shed light on the interpreting aptitude test and interpreting pedagogy.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Li Dechao, who offered me constructive guidance and enthusiastic encouragement during my Ph.D. study. From him, I learned to think critically and be open-minded in my research. I sincerely thank him for making my study in Hong Kong well worth it. My thanks are also extended to many of the academic staff in the Department of Chinese and Bilingual Studies of the Hong Kong Polytechnic University, who broadened my view of research and had discussions with me about my dissertation. I would also like to thank my committee members, Prof. Minhua Liu and Prof. Ren Wen, for their valuable insights.

I am grateful to Prof. Franz Pöchhacker, Prof. Arnt Lykke Jakobsen, Prof. Daniel Gile, Prof. Haidee Kruger and Prof. Heidi Salaets, who offered invaluable suggestions regarding my research at the 2018 CETRA summer school, which benefited me enormously. The enlightening discussions with them solved daunting questions of mine, broadened my horizons, and revealed new directions for my research.

Special appreciation goes to Prof. Norman Segalowitz of Concordia University in Canada for his kind response to my questions during our email communication, which cleared my doubts about important issues in my research. His studies have served as valuable references for mine. My thanks also go to Dr. Rebecca Tipton, whose lectures further broadened my view in interpreting research during my stay for the attachment program at the University of Manchester in 2019.

I am also indebted to everyone who has been with me in AG 518 in the Hong Kong Polytechnic University since 2016. They had discussions with me about my concerns and gave me valuable suggestions at the planning and development stages of this research. Their company, support and friendship will be most preciously remembered by me. My thanks also go to the 28 participants who took part in my experiment and Joe, Mr. CHAN Kam Yiu, for his kind technical assistance during the data collection process.

Finally, my gratitude to my family can never be fully expressed. They have sacrificed a lot to make my academic pursuit possible. This dissertation would not have been accomplished without their unwavering understanding and encouragement.

Table of contents

List of abbreviations

SI: simultaneous interpreting

Cognitive fluency related:

LA: lexical access

LR: lexical retrieval

AC: attention control

SS: speaking span

CV: coefficient of variance

Utterance fluency indicators:

SR: speech rate

AR: articulation rate

MLR: mean length of run

PTR: phonation time ratio

SP mean: mean number of silent pauses per minute

SP length: mean duration of silent pauses

FP mean: mean number of filled pauses per minute

FP length: mean duration of filled pauses

REPAIRs: repairs, repetitions and false starts

REPAIR mean: mean number of REPAIRs per minute

REPAIR length: mean duration of REPAIRs

List of tables

List of figures

List of charts

List of appendices

Chapter 1 Introduction

1.1 Research motivation

Interpreting quality is "an essentially relative and multi-dimensional concept" and a complex theme in which different aspects of the product and performance of interpreters play a part (Pöchhacker, 2004, p. 153). Fluency is regarded as one of the most important quality criteria in interpreter training institutions (Liu, Zhang, & Wu, 2008), distinct from other criteria such as fidelity (message accuracy and completeness). As a sub-parameter of interpreting quality, fluency plays a non-negligible role in the overall judgment of interpreting output. Although it ultimately provides no guarantee of the reliability of the interpreter, the role of fluency in the overall impact of the target speech should not be neglected and it is an important feature of successful interpreting (Mead, 2000, 2005). Fluency could be the single aspect of an interpretation which distinguishes the performance of students from that of professional interpreters (Altman, 1994). Moreover, fluency can reflect the ease of cognitive processing of interpreters. However, it is also argued in previous studies that the significance of fluency has been underestimated, with the tendency of interpreters and interpreting service users to prioritise accuracy or fidelity over fluency (Bühler, 1986; Chiaro & Nocella, 2004; Kurz, 1993, 2001, 2003; Moser, 1996; Yu & van Heuven, 2017).

Only recently has fluency been regarded as an individual component of interpreting quality by interpreting scholars in their research investigating the constituents of fluency and its impact on user perception (Rennert, 2010). Previous studies on fluency in interpreting are mostly descriptive, focusing primarily on disfluencies, which signify difficulties and uncertainties in the cognitive processes of interpreting (e.g. Bakti, 2009; Cecot, 2001; Tissi, 2000). Parameters of fluency have been included in most previous explorations of interpreting quality, but the constructs of fluency are often not operationalised and tend to be "merely labeled but not described", making rater bias inevitable (Liu, 2015a, p. 20). Links between objective measures of performance fluency in interpreting and overall ratings of interpreting performance have been found in both simultaneous (Han, 2015a) and consecutive (Yu & van Heuven, 2017) interpreting. However, with scanty empirical evidence, the cognitive bases of fluency in interpreting output are still ambiguous, though interpreting is generally acknowledged as a highly demanding cognitive task. Cognitive science is "the interdisciplinary attempt to understand the mind", involving four branches (i.e., behavioural and brain sciences, formal disciplines, social sciences and the philosophy of mind). The hallmark of cognitive science is that it draws on methods and results of all these branches to give a global understanding of the mind (Stainton, 2006, p. viii). The multidisciplinarity afforded by a cognitive science perspective is indispensable for studies on fluency since it is a multidimensional construct (Segalowitz, 2010). Moreover, there is little longitudinal research investigating the relationship between different dimensions of fluency in interpreting, in particular at different developmental stages of interpreting expertise.

This research is a multi-dimensional and longitudinal exploration of fluency in trainee interpreters' L2 (English) to L1 (Chinese) simultaneous interpreting (SI). The relationships between three dimensions of fluency in SI (i.e., cognitive fluency, utterance fluency and perceived fluency) (Segalowitz, 2010) are explored. The role of cognitive fluency in the development of SI utterance fluency is investigated to explore the cognitive bases of SI fluency. The role of SI training and cognitive load on trainee interpreters' SI utterance fluency are explored to make a multifaceted analysis of the role of cognitive fluency factors. The relationship between objective indicators of utterance fluency in trainee interpreters' SI and fluency ratings is included in the investigation so as to form a full picture of different domains of fluency in interpreting.

1.2 Definitions and constructs of fluency

1.2.1 Defining fluency

Fluency is a multi-level, multi-dimensional construct and there are no broadly accepted definitions of fluency in previous research on either interpreting or English as a native or second language. Scholars have defined different connotations and denotations of fluency, which involve fluidity, acceptability, cohesion and so on (e.g. Fillmore, 2000; Guillot, 1999; Lennon, 1990; Pradas Macías, 2006; Schmidt, 1992).

Some scholars regard fluency as a kind of language competence from the perspective of generative linguistics. Leeson (1975, p. 136) defined fluency as "the ability of the speaker to produce indefinitely many sentences conforming to the phonological, syntactical and semantic exigencies of a given natural language on the basis of a finite exposure to a finite corpus of that language". However, he was concerned primarily with phonology and the acquisition of generative rules and took no account of situational interaction and the way language was actually learned (Brumfit, 1984; Zhang, 1999a). Fillmore (2000, p. 51) identified four aspects of fluency from the perspective of language production to judge whether speakers were fluent in their native language: 1) the ability to "talk at length with few pauses, the ability to fill time with talk"; 2) the ability to "talk in coherent, reasoned, and 'semantically dense' sentences" and the tendency "not to fill discourse with lots of semantically empty material"; 3) the ability to "have appropriate things to say in a wide range of contexts"; and 4) the ability to "be creative and imaginative in their language use", to express ideas in novel ways. Fillmore's (2000) categories linked the language system with personality characteristics and drew attention to the interaction between language and world knowledge (Brumfit, 1984).

In terms of L2 fluency, representative definitions of fluency include those of Lennon (1990) and Schmidt (1991). Lennon (1990) distinguished fluency in the broad sense from that in the narrow sense. In the broad sense, it serves as a term for oral proficiency and represents "the highest point on a scale that measures spoken command of a foreign language" (p. 389), while in the narrow sense, it refers to one component of oral proficiency. A working definition of fluency given by Lennon (2000, p. 26) was "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing". Fluency in the narrow sense is distinct from components like grammatical knowledge and vocabulary size (Pinget, Bosker, Quené, & De Jong, 2014) and was defined as "the ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation" (Housen, Kuiken, & Vedder, 2012, p. 2).

Some scholars proposed to add a standard of acceptability when defining fluency (Guillot, 1999; Meisel, 1987; Sajavaara, 1987). It was pointed out by Lennon (1990) that fluency was a performance phenomenon and it was "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" (p. 391). Schmidt (1991) stated that fluency is an automatic procedural skill not requiring much effort, in contrast with non-fluent speech, which is effortful, which was generally in line with Lennon's (1990) view. Besides, the feature of coherence was emphasised by Zhang (1999), who defined L2 oral fluency as the ability to express thoughts in a smooth, coherent and acceptable way in the second language.

Classifications of sub-parameters of fluency are not homogeneous in previous research. Three principal categories were identified (i.e., hesitations, corrections and speech rate), in Riggenbach's (1991) microanalysis of non-native speakers' conversation. Silent pauses and filled pauses were included in the category of hesitations, and repetitions and false starts were included under the category of corrections. Guillot (1999) proposed two perspectives of fluency, i.e., "a sense of ease and the absence of (obvious) hesitations" and "coherence, effectiveness and intelligibility" (cited in Pradas Macías, 2006, p. 27). The three dimensions of fluency proposed by Skehan (2003) and Tavakoli and Skehan (2005) were speed fluency (relevant to speech rate), breakdown fluency (hesitation phenomena such as the number and length of pauses), and repair fluency (false starts, repairs, repetitions etc.), which focused on the utterance perspective of fluency. Their categorisation of fluency dimensions is adopted in the current research.

In terms of functions of fluency, each type of fluency indicator can be linked to a certain stage in the process of speech production, from conceptual planning and grammatical encoding to articulatory planning and monitoring (Bakti, 2009; Gósy, 2007; Levelt, 1983; Schnadt & Corley, 2006). Definitions of different types of fluency indicators and their functions are elaborated below. Pauses, which are further categorised into filled and unfilled (silent) pauses, are periods during which no acoustic signal occurs for at least 200-270ms (Hargrove & McGarr, 1994). The duration of a filled or unfilled pause indicates planning processes (Dechert & Raupach, 1980). Unfilled pauses refer to silent periods between vocalisations, and filled pauses to interruptions of speech flow by non-lexical sounds such as “ah”, “er” or “erm” (Cenoz, 1998). Unfilled pauses have multiple functions in language production, which may be grammatical, communicative or hesitative (Simone, 1995). Filled pauses may reflect brief attention to planning or retrieval and include fillers or the lengthening of syllables or words. A false start is an utterance that is begun and then in some way abandoned or reformulated. A repetition occurs when previously produced speech is produced again, which is a device that may be used to allow time for on-line planning (Foster, Tonkyn, & Wigglesworth, 2000). It should be noted that disfluent repetitions differ from repetitions used for rhetorical effects. Repetitions under the domain of repair fluency in the current research only refer to disfluent ones. A repair is a self-correction of a speaker when he or she identifies an error during or immediately after production and stops to reformulate the produced speech (Foster et al., 2000). Repairs reflect the monitoring mechanism used to verify the correctness of motor activity and response output (Postma, 2000).

Following previous definitions, we propose a tentative working definition of fluency in interpreting which encompasses the three domains of fluency reviewed above:

Fluency in interpreting refers to the smooth, clear, efficient and intelligible oral transfer of the original message into the target language under temporal and cognitive constraints, with reasonable pausing and/or hesitations and leaving the listeners with a sense of ease.

1.2.2 Conceptualising cognitive fluency

Three domains of fluency are labelled by Segalowitz (2010): cognitive fluency, utterance fluency and perceived fluency, which provide rich connotations of fluency with the cognitive mechanism included. This framework, which focuses on the cognitive processing that underlies fluency and on the influence of social context, provides a perspective of dynamic systems for fluency and its development (Segalowitz, 2016).

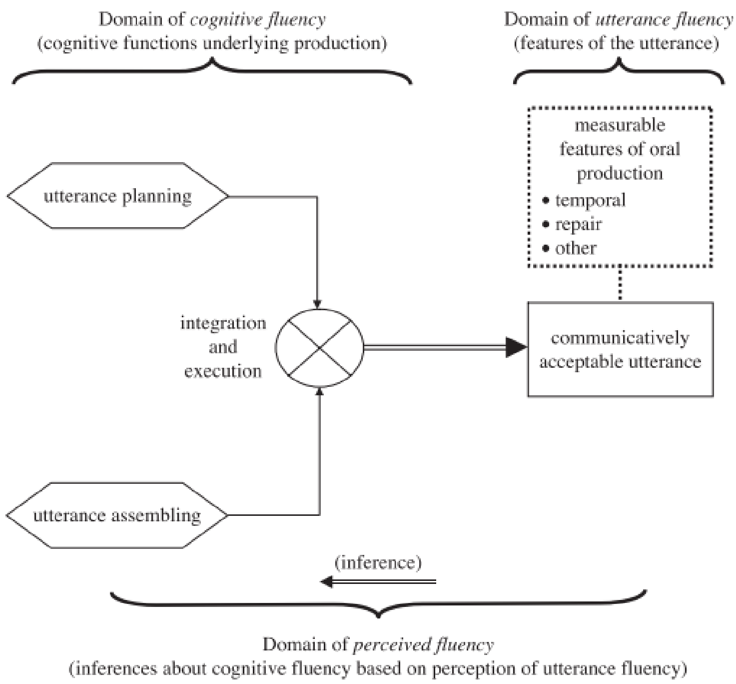


Figure 1.1 Three domains of fluency and their relationships (Segalowitz, 2010)

Cognitive fluency refers to the fluid operation (speed and efficiency) of the speaker in their efforts to "mobilize and integrate the underlying cognitive processes responsible for producing utterances with the characteristics that they have" (Segalowitz, 2010, p. 48). The ability of the cognitive system per se reflects how efficient and fluid the mobilisation and integration of processes can be (Segalowitz, 2010). As shown in

Figure 1.1, the underlying cognitive system carries out functions of utterance planning and assembling. With the integration of these functions, utterances are executed with the desired features of oral production. The domain of cognitive fluency is the operation of these planning and assembling functions and their integration and execution (Segalowitz, 2010). Utterance fluency, which refers to the set of objectively determined oral features of utterances, reflects the impact of the underlying cognitive processes and represents the characteristics a speech sample possesses, e.g. the temporal, hesitation and repair features. Perceived fluency refers to "the inferences listeners make about a speaker's cognitive fluency based on their perception of the utterance fluency" of the speech output (Segalowitz, 2010, p. 48).

The above conceptualisation of cognitive fluency is consistent with older understandings of the cognitive bases of language production. It is generally accepted that automatisation is required for fluency, as it enhances both the fluidity and the efficiency of the speaker's underlying cognitive processing (Anderson, 1983b). According to Goldman-Eisler (1968, p. 6), the speech act is "a dynamic process, demanding the mobilisation in proper sequence of a series of complex procedures and is the temporal integration of serial phenomena". She pointed out the central role of cognitive mechanisms in forming the temporal phenomena of oral fluency. Cognitive ability permits inductive learning of generative rules which underlie high-level linguistic competence and forms the two bases of fluent language production together with physiological characteristics (Leeson, 1975). It was also pointed out by Rehbein (1987) that the rapid speed, automaticity and efficiency of underlying mechanisms were responsible for the fluency of L2 speech.

Constructs of cognitive fluency include both linguistic knowledge and skills and nonlinguistic features such as working memory capacity and conceptualising skills (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013). Cognitive processes involved in cognitive fluency include the speed and efficiency of semantic retrieval, linguistic attention control with the functions of languages-directed attention linked to L2 grammatical aspects, and operations in working memory, among others (Segalowitz, 2010, 2016).

Candidate measures of cognitive fluency most explored in previous research include the speed and efficiency of lexical access and flexibility in the control of linguistic attention. Experimental tasks used to explore semantic retrieval include semantic classification task, sentence construction task, picture-naming task and so on (De Jong et al., 2013; García-Amaya, 2012; Lim & Godfroid, 2014; Seeber & Freed, 2004; Segalowitz & Frenkiel-Fishman, 2005). The flexibility of linguistic attention is related to the way grammatical elements direct attention to relationships between utterance elements and can be operationalised as a switch cost elicited from an alternating runs experimental design (Rogers & Monsell, 1995). The switch cost, also referred to as shift cost, represents the burden on the cognitive processing system of having to change the focus of attention (Segalowitz, 2010).

Linguistic attention is also called language-directed attention, which forms an essential component of cognitive fluency (Segalowitz, 2010; Talmy, 2008). For language-directed attention, "the control of attention originates from the linguistic message itself and is directed back to the mental representation that is associated with the meaning of the message" (Segalowitz, 2010, p. 95). Language-directed attention is associated with the grammatical features of language. "Grammatical devices direct the recipient's attention by modulating the focus and saliency of elements in the linguistic message and by directing the building of mental spaces or representations of the meaning of the message." (ibid., p. 99). Language-directed attention contrasts with attention-to-language, which appears not different from general attention-to-the-world, where "control of attention originates from outside the message and is directed toward the surface level of the message" (ibid., p. 95). The control of attention is operationalised as a person's ability to "shift focus of attention from one language-based attention-directing function to another" and a superior ability to make these shifts rapidly is assumed to indicate better control of language-directed attention (Segalowitz & Frenkiel-Fishman, 2005, p. 646).

Research on cognitive fluency in L2 sheds light on the exploration of fluency in interpreting, a cognitively demanding language processing task. Since interpreting involves the conversion of message from the source to target language, the investigation into cognitive fluency in the current research includes both L1 and L2 measures and takes lexical retrieval, which involves an oral word translation task, into account. Based on the conceptualisation of cognitive fluency in L2 discussed above, a tentative working definition of cognitive fluency in interpreting is proposed below:

Cognitive fluency in interpreting refers to the efficient mobilisation of cognitive processes involved in interpreting, which includes the mobilisation, integration and coordination of mental processes that are responsible for the fluent interpreted output. These processes include the speed and efficiency of lexical access and lexical retrieval, linguistic attention control, working memory operations and so on.

1.3 Research questions

Following the domains of fluency proposed by Segalowitz (2010), this research offers a multidimensional exploration of fluency in English (L2) to Chinese (L1) SI of trainee interpreters and explores the relationships between three domains of fluency in interpreting (i.e., cognitive, utterance and perceived fluency). Cognitive fluency is explored from the perspectives of the efficiency of lexical access, lexical retrieval, linguistic attention control and working memory capacity, elicited through a semantic classification task, a word translation task, a category judgment task and a speaking span task. Measures of utterance fluency in SI performance were obtained through simulated SI tasks at two time slots of interpreting training (i.e., the beginning and the end of an SI training period of one academic term). For utterance fluency, the three categories of operationalised measures are speed fluency (speech rate relevant), breakdown fluency (hesitation phenomena such as the number and length of pauses), and repair fluency (false starts, corrections, repetitions etc.) (Skehan, 2003; Tavakoli & Skehan, 2005). Human raters with professional interpreting experience were invited to assess fluency in SI performance as a measure for perceived fluency. The overall framework of the constructs of fluency in SI adopted in the current research is listed in Table 1.1.

Table 1.1 Constructs of fluency in SI in the current research

	Dimensions	Operationalised
Fluency in SI	Cognitive fluency	Lexical access
		Lexical retrieval
		Linguistic attention control
		Working memory capacity
	Utterance fluency	Speed fluency
		Breakdown fluency
		Repair fluency
	Perceived fluency	Ratings

Main research questions (RQ) of the current research are as follows:

RQ 1 Can cognitive fluency measures of trainee interpreters, operationalised as measures of lexical access, lexical retrieval, linguistic attention control and working memory capacity, predict the development of trainee interpreters' SI utterance fluency under conditions of low and high cognitive load respectively?

RQ 1.1 To what extent can gains in SI utterance fluency be predicted by cognitive fluency measures of trainee interpreters under conditions of low and high input rate, respectively?

RQ 1.2 Does the inclusion of lexical retrieval and working memory capacity measures lead to higher predicting power of cognitive fluency measures for the development of utterance fluency in trainee interpreters' SI output?

Hypotheses: It is hypothesised that better cognitive fluency abilities (more efficient lexical access and lexical retrieval, better linguistic attention control and larger working memory capacity) can to some extent predict the development of trainee interpreters'

SI utterance fluency, operationalised as measures of speed, breakdown and repair fluency in the post-test partialling out correspondent measures in the pre-test. It is also hypothesised that cognitive fluency measures have a higher predicting power for the development of trainee interpreters' SI utterance fluency under conditions of high cognitive load than under low cognitive load conditions. In addition, the inclusion of measures of lexical retrieval and working memory capacity is hypothesised to lead to a higher predicting power of cognitive fluency measures for the utterance fluency development in trainee interpreters' SI output.

RQ 2 How do trainee interpreters' cognitive fluency, SI training and input rate of source speech influence their utterance fluency in SI output, comparing measures of speed, breakdown and repair fluency at two time slots of SI training?

RQ 2.1 Is there any interplay between cognitive fluency measures, SI training and input rate in terms of their influence on the utterance fluency of trainee interpreters' SI output?

RQ 2.2 How does cognitive fluency influence measures of utterance fluency in trainee interpreters' SI performance?

RQ 2.3 How does SI training influence measures of utterance fluency in trainee interpreters' SI performance?

RQ 2.4 How does input rate influence measures of utterance fluency in trainee interpreters' SI performance?

Hypotheses: It is hypothesised that cognitive fluency measures, SI training and input rate have a significant impact on the utterance fluency of trainee interpreters' SI output. Significant differences in one or more utterance fluency indicators are expected between one or more low and high cognitive fluency groups of trainee interpreters, between pre-training and post-training SI performance and between measures under low and high input rate conditions. Some interaction effects among cognitive fluency, SI training and input rate are expected.

RQ 3 What is the relationship between objective indicators of utterance fluency in trainee interpreters' SI output and perceived fluency, as assessed by human raters?

RQ 3.1 Which indicators of utterance fluency (speed, breakdown and repair fluency) are most related to perceived fluency, as assessed by raters?

RQ 3.2 To what extent can SI utterance fluency indicators predict perceived fluency as rated by human raters?

Hypotheses: It is hypothesised that some of the measures of SI utterance fluency correlated significantly with perceived fluency and SI utterance fluency indicators can to some extent predict perceived fluency, as assessed by raters with professional interpreting experience.

1.4 Structure of the dissertation

This dissertation consists of eight chapters.

Chapter One sets out to introduce the research motivation of the current research and defines the concepts and constructs of fluency. The connotations of cognitive fluency are explained within the framework of three domains of fluency (Segalowitz, 2010). Research questions are raised, after which the structure of the dissertation is described.

Chapter Two reviews previous studies on fluency in second language learning and interpreting, including studies on the relationship between cognitive fluency and utterance fluency and between utterance fluency and perceived fluency in second language learning. The reviewed studies offer important methodological references for the current research. Previous studies on cognitive factors, utterance fluency and the perception of fluency in interpreting are reviewed and main factors affecting fluency in interpreting are discussed, which helps to identify the research gap and serves as a foundation for the current research.

Chapter Three expounds relevant theoretical underpinnings, including the Effort Model of SI, the bilingual speech production model and the embedded-processes model of working memory.

Chapter Four explains the research design. Operationalisations of cognitive fluency through behavioural experiments (i.e., the semantic classification task, word translation task, category judgment task and speaking span task) are explained. Measurement of utterance fluency indicators and the rating method of perceived fluency are illustrated. Moreover, this chapter elaborates the data collection process, including the recruitment of participants, speech manipulation for the SI tasks and the procedures of the behavioural experiments and SI tasks. The annotation system of the interpreted output is also explained.

Chapter Five explores the role of cognitive fluency measures in the development of utterance fluency in trainee interpreters' SI output in response to RQ 1. A series of multiple linear regression analyses are performed to investigate the predicting power of cognitive fluency on the development of each utterance fluency indicator under both low and high input rate conditions. Results of the analyses are discussed to explore the influence of each cognitive fluency measure (i.e., lexical access, lexical retrieval, linguistic attention control and working memory capacity) on gains in utterance fluency. It also discusses the added explanatory power of lexical retrieval and working memory capacity for utterance fluency development, compared to the model that includes only lexical access and linguistic attention control.

Chapter Six investigates the effects of cognitive fluency, SI training and input rate on utterance fluency indicators in trainee interpreters' SI output. Repeated measures ANOVA analyses are conducted to explore the individual main effects and interplay of cognitive fluency, SI training and input rate on each utterance fluency indicator. Findings are summarised and discussed to see how utterance fluency measures of trainee interpreters' SI output are influenced by these factors.

Chapter Seven taps into the relationship between utterance fluency measures and perceived fluency, as assessed by raters with professional interpreting experience. The predicting power of utterance fluency indicators in trainee interpreters' SI output for perceived fluency is also explored.

Chapter Eight is the conclusion of the current research, which summarises the main findings, discusses the significance and implications of the research, and reflects on the study's limitations and directions for future research.

Chapter 2 Literature review

This chapter first reviews previous studies on the relationship between cognitive fluency and utterance fluency and between utterance fluency and perceived fluency in second language learning, offering theoretical and methodological references for the current research. Previous studies on cognitive factors, utterance fluency and the perception of fluency in interpreting are also reviewed to identify the research gap and serve as a foundation for the current research. Main factors affecting fluency in interpreting are discussed, identifying the variables of training and the input rate, which are manipulated in this research.

2.1 Fluency in second language learning

2.1.1 Relationship between cognitive fluency and utterance fluency

The relationship between cognitive fluency and utterance fluency has been comparatively under-explored in past research. This line of research investigates which aspects of utterance fluency could reflect the ease and efficiency of L2 speaking production (cognitive fluency) and which features of cognitive fluency can predict measures of utterance fluency. Fluidity is the predominant feature of fluency in the minds of researchers, but fluidity itself is a multidimensional construct and the perspective of cognitive science is indispensable (Segalowitz, 2010) for understanding it. De Jong et al. (2013) pointed out that cognitive fluency involves both linguistic knowledge and skills and nonlinguistic features like working memory and conceptualising skills, a point generally in line with Lennon's (2000) claim that fluent performance is the outcome of the speaker's adequate linguistic and pragmatic knowledge and sufficient processing abilities. According to Goldman-Eisler (1968), the crucial factor in fluency is the efficiency of the speaker's cognitions underlying speech production. Inefficient linking of words to meanings might slow down the overall processing and create overload problems in short-term memory, requiring the speaker to reduce speech rate and potentially being responsible for some speech hesitations and pauses.

It is generally accepted that automatisation is required for utterance fluency, as automatisation enhances both the fluidity and the efficiency of the speaker's underlying cognitive processing. Automatic processing was defined as ballistic or unstoppable processing by Favreau and Segalowitz (1983), who indicated that while speed of processing is not diagnostic of fluency differences, ballistic processing is. Segalowitz and Segalowitz (1993) argued that automaticity could be studied by distinguishing processing speed from processing stability and proposed to examine processing efficiency by considering both speed (reaction time) and intraindividual variability of reaction times. One set of reaction times might be faster than another for either of two reasons: the speedup of underlying mechanisms responsible for the faster set, and restructuring or higher organisational efficiency of the action carried (Cheng, 1985; Segalowitz, 2010). It is argued that greater organisational efficiency is achieved through both faster performance and greater processing stability. The latter can be operationalised as CV (coefficient of variance) of reaction time. CV was the standard deviation of an individual's reaction times divided by that person's mean reaction time, which provided "a measure of response-time variability adjusted for overall speed of response" (Segalowitz & Freed, p. 176). CV can be interpreted as a measure of automatisation. A change in CV implies restructuring of underlying processes (a qualitative increase in processing efficiency), which differs from a change in reaction time due to simple speed-up of those (Segalowitz & Segalowitz, 1993; Segalowitz, Seeber, & Wood, 1998).

Previous studies on the relationship between cognitive fluency and utterance fluency in second language learning are reviewed and summarised in the following section.

2.1.1.1 Longitudinal studies

One line of previous research investigates the development of utterance fluency within speakers over time. A development in one or more measures of utterance fluency over time is assumed to be traced to the development of cognitive fluency (De Jong et al., 2013). Some longitudinal studies were conducted to investigate the development of utterance fluency in L2. Towell, Hawkins, & Bazergui (1996) conducted longitudinal studies following L2 learners of French over three years, attempting to explicitly explore processes and mechanisms underlying oral fluency. Gains in fluency were found with respect to mean length of run (MLR) and speech rate (SR), but not to mean duration of silent pauses. The gains were attributed to the proceduralisation of knowledge, including the knowledge of syntax and lexical phrases. Their research further indicated that gains in fluency reflected complex underlying changes in speech generation, and gains in smoothness resulted from the restructuring of underlying process. They adopted the perspective of ACT theory (Anderson, 1983a), that skilled performance involves the conversion of declarative knowledge into procedural knowledge. Towell (2002) examined how and why oral fluency developed at different rates among undergraduate learners of French. Qualitative analysis found that gains in temporal variable measures of some learners were made by modifying the pausing behaviour and making the syntax more complex. Towell's (2002) research pointed out that individual factors like working memory were believed to be important for fluency. The longitudinal research conducted by Derwing, Munro, Thomson, & Rossiter (2009) compared L1 and L2 English fluency three times over two years in adult immigrant speakers of Slavic and Mandarin. A relationship between L1 and L2 fluency was indicated by ratings of trained judges in the initial stages of L2 exposure, leading the researchers to believe that the close relationship between L1 and L2 temporal characteristics suggests that fluency is governed by an underlying trait. Pauses per second, speech rate, and pruned syllables per second were found to be related to listeners' judgment in both languages, but vowel durations were not.

2.1.1.2 Correlational research

Another line of previous research directly relates measures of cognitive fluency to utterance fluency measures and explores their relationships (De Jong et al., 2013). This line of research is also concerned with what factors of cognitive or utterance fluency can best predict overall L2 fluency.

In the first study directly relating utterance fluency to cognitive fluency, Segalowitz and Freed (2004) investigated the role of learning contexts in L2 acquisition. Main variables selected for utterance fluency, among others, were speech rate, mean length of run excluding silent pauses or hesitations longer than 400ms, mean length of run excluding filled pauses, and longest run excluding silent or filled pauses. The cognitive variables measured in this study were the speed and efficiency of lexical access and linguistic attention control. The speed of processing was indexed by reaction time and the efficiency of processing by the CV of reaction time. It was found that cognitive abilities of the speed and efficiency of lexical access were positively related to oral fluency in their pre-tests. No significant correlations were found between attention control speed and oral fluency. The efficiency of linguistic attention control was found to have a negative relationship with speech rate in the post-tests, which might suggest that the greater efficiency of learners in attention shifting led to overall slower speech rate. It was concluded that the speed and efficiency of L2-specific cognitive processing were implied in oral performance. In this study, L1 performance on the same tasks was partialled out, so the task results in this study did not reflect subjects' general cognitive abilities but rather their L2-specific cognitive skills. However, et al.'s (2014) study indicated that partialling out effects of L1 from L2 performance is not necessarily better than traditional measures of L2 in predicting perceived fluency, though the residuals did predict the variance in fluency ratings to a large extent. This might be attributed to the fact that the raters had no access to L1 fluency measures of these speakers, thus partialling out L1 fluency did not lead to a higher predictive power of fluency ratings.

De Jong et al. (2013) investigated which aspects of L2 utterance fluency were indicators of L2 cognitive fluency and explored to what extent objectively measured aspects of utterance fluency could be explained by measures of L2 linguistic knowledge and processing skills which underlay L2 cognitive fluency. Three knowledge tests were used to measure knowledge of vocabulary, grammar, and pronunciation. Several processing tests included a picture naming task to tap lexical retrieval speed, a sentence completion task to test the speed of morphosyntactic knowledge access, and a delayed picture naming task to measure the speed to articulate speech plans. The measures for utterance fluency included those for breakdown fluency (number and mean duration of silent pauses, and number of filled pauses), speed fluency (inverse articulation rate, i.e. speaking time divided by the total number of syllables) and repair fluency (number of corrections and number of repetitions). The research results showed that linguistic skills were related to average syllable duration most strongly, which explained 50% of the individual variance; and the average pausing duration was only weakly related to linguistic knowledge and processing skills. Overall, all measures of utterance fluency were related to one or more measures of underlying cognitive fluency. This research differs from other similar research in that it measured pausing and speed separately by the measurement of mean syllable duration (inverse articulation rate excluding pauses) while most studies used a measure that incorporated both pausing and speed of speech, i.e. speech rate calculated as syllables divided by the total time, which incorporates both pausing and speech speed. The research of De Jong et al. (2013) differs from that of Segalowitz and Freed (2004) in that, in De Jong et al.'s (2013) study, L2 performance on linguistic knowledge and skills was measured without partialling out the performance on similar L1 performance. While acknowledging that part of the variance of L2 linguistic skill measures must be language-independent, they emphasised that only L2 linguistic knowledge and skills were investigated in this study. It was fully acknowledged that cognitive fluency included nonlinguistic features like working memory capacity and conceptualising skills.

The predictive power of L2 cognitive fluency for L2 oral competence was also explored in the study conducted by Hu and Wang (2017), with reference to the predictive power of L2 oral fluency for L2 oral competence. Two cognitive indicators were reaction time of sentence construction task and attention shifting cost. SR (speech rate) and MLR (mean length of run) served as indicators of utterance fluency. Results of hierarchical regression analysis indicate that cognitive fluency significantly improved the predictive power for L2 oral competence and bore a much higher degree of contribution to the prediction than did performance fluency. This study sheds light on the relationship between cognitive fluency and overall oral proficiency.

2.1.1.3 A special focus on memory

Exploration of the relationship between cognitive and utterance fluency is closely related to memory. Working memory refers to "a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning" (Baddeley, 1992, p.556). Working memory requires processing capacity and has a limited storage capacity. It plays an essential role in both L1 and L2 language production (Daneman, 1991; Fortkamp & Bergsleithner, 2007; Guará-Tavares, 2013; O'Brien, Segalowitz, Freed, & Collentine, 2007; Wen, 2012).

Working memory is an internal cognitive factor affecting both L1 and L2 language processing, though findings of whether memory plays an independent role in language production are not consistent. Research through a speaking span test by Daneman (1991) showed that the capacity of working memory was correlated to oral fluency and individuals with small speaking span were less fluent and tended to make more speech errors. Following this study, Fortkamp (1999) explored the correlation between working memory capacity and L2 oral fluency. The study found that there was significant correlation only between a reading-related task and working memory capacity, as measured through the speaking span test and reading span test in both L1 and L2. The correlation between working memory and the development of L2 fluency was also explored by Mizera (2006), but there was a lack of correlation between working memory and metrics of L2 speech performance. Moreover, working memory capacity of learners was found to be positively related to not only fluency but also accuracy and complexity in L2 oral performance (Fortkamp & Bergsleithner, 2007; Fortkamp, 2000). Speaking span is regarded as a significant predictor of L2 fluency, accuracy and complexity, which also partially explains the variation of L2 oral performance (Fortkamp, 2000; Guará-Tavares, 2013).

Kroll, Michael, Tokowicz, & Dufour (2002) compared the performance of more and less fluent L2 speaker groups in a word-naming task and a translation task. The study found that higher span participants were slower than lower span participants at translating cognates in both directions, while were faster when translating noncognate words. It was explained that higher span leaners, though additional memory resources enabled them to avoid false cognates, suffered a processing cost in the task. The overall results suggest that working memory resources affect translation performance. The research of O'Brien et al. (2007) investigated the relationship between phonological memory and gains in L2 fluency of native English speaking adults in two learning contexts, and was the first research to demonstrate a relationship between phonological memory and adults' L2 oral fluency. L2 oral fluency was measured by temporal and hesitation phenomena and phonological memory was operationalised as performance in serial nonword recognition, which required participants to decide whether two strings of nonwords were presented in the same order or not. The main finding of the research was that phonological memory was related to L2 oral fluency development of adults over time, though the performance of serial nonword recognition itself did not change over the same period. Therefore, phonological memory seems to be an important skill for L2 speech production development in adults. Sunderman and Kroll (2009) investigated whether internal cognitive resources was to a certain extent necessary for individuals to take full advantage of study-abroad experiences. It was found that both internal and external resources (memory and context) made independent contributions to processing speed and accuracy and high span learners were faster and more accurate than their low span counterparts regardless of learning context. However, working memory resources alone were not sufficient to affect L2 production, which differed from O'Brien et al.'s (2007) finding that phonological memory affected L2 oral fluency regardless of context. Similarly, Mizera (2006) failed to find a significant correlation between phonological memory (measured by nonword recognition task) and L2 translation tasks or oral production measures in low and high L2 proficiency adults. Wen (2012) found that better phonological working memory was an advantage for speed fluency measures like speech rate and replacement while executive working memory seemed to affect measures of repair fluency like false start and reformulation to a larger extent.

As can be seen from the review above, previous studies have investigated the underlying role of cognitive fluency in utterance fluency and explored the relationship between cognitive fluency and utterance fluency in second language learning from multiple angles, which provides valuable methodological and theoretical references for the current research.

2.1.2 Relationship between utterance fluency and perceived fluency

Previous studies on measures of utterance fluency often concerned the relationship between objective measures of fluency and the perception of fluency. Scholars have explored which aspects of utterance fluency would affect listeners' perception of fluency, which is important for exploring the possibility of automatic assessment of fluency and language pedagogy.

A number of quantitative variables related to perceived fluency have been identified in studies of temporal features in L1 and L2 speech (De Jong et al., 2013; Freed, 1995; Goldman-Eisler, 1968; Grosjean, 1980; Lennon, 1990; Nation, 1989; Riggenbach, 1991; Towell et al., 1996), such as speech rate, number of silent and filled pauses, and other hesitations such as repetitions and repairs. In Lennon's (1990) study, ten native-speaker teachers of EFL rated the recordings of four advanced EFL learners. Two isolatable components of fluency were identified: temporal components (e.g. pruned words per minute, the percent of T-units followed by pause) and vocal disfluency marker components such as filled pauses and repetitions, but not necessarily self-corrections, which appeared to be a poor fluency indicator. Riggenbach (1991) related ratings by English instructors to temporal measures of speaking performances in her investigation of spontaneous speech in dialogues of six learners of L2 English. Hesitation phenomena (unfilled and filled pauses), repair phenomena (repetitions, false starts) and the rate and amount of speech were analysed. It was found that the ratings were related primarily to speech rate and silent pauses; repair phenomena seemed to play a less evident role in fluency. Kormos (2006) outlined the ten most frequently used temporal variables of fluency in previous research and pointed out that speech rate (number of syllables per minute), mean length of run (number of syllables in utterances between pauses of minimum 0.25s) and phonation time ratio (PTR, percentage of speaking time as a proportion of the overall speech time) were the best predictors of fluency according to most studies. Automated measures of oral fluency (speech rate, measures of silence) were obtained by De Jong, Schoonen and Hulstijn (2009), using eight speaking tasks employing pictures depicting short scenarios to elicit speech samples in both L1 and L2 from the participants. Significant relationships between L1 (English) and L2 (Dutch) performance were found. Their research suggested that, using L1 as a baseline, the oral variables best indicating L2 fluency were effect sizes for L1-L2 differences on the percent of silent pauses per word (not including the length of silent pauses), words per second speech rate especially excluding filled pauses, and percentage of corrections or self-repairs per word. Bosker, Pinget, Quené, Sanders, & De Jong (2012) investigated the contributions of three dimensions of fluency, i.e. pauses, speed and repairs to perceived fluency. Results of linear regression analyses revealed that the best predictors for the subjective fluency ratings were pause and speed measures and that the contribution of repair measures was only very small.

Some of the explorations of temporal variables related to perceived fluency took task types into consideration. The research of Cucchiarini, Strik, & Boves (2002) explored the relationship between objective measures of speech and perceived fluency in both read and spontaneous speech of non-native Dutch speakers. The speech fragments were scored by human raters and analysed by a speech recogniser for the calculation of several objective measures of speech quality. The research showed that objective measures could be employed to predict fluency ratings. The correlations between speed-related measures (SR, AR, MLR and PTR) and silent pause measures (duration and number per minute) and fluency ratings were significant. The predictive power of such measures was stronger for read speech than for spontaneous speech. The adequacy of the employed variables seemed to depend on the speech material and the specific task performed. Derwing, Rossiter, Munro, & Thomson (2004) investigated whether temporal measures of fluency were related to untrained raters' assessment of fluency in low-proficiency L2 speech and whether they varied across tasks. Samples from the speech of 20 beginner Mandarin learners of English were collected to be rated by untrained judges. It was found that pausing and pruned speech rate (excluding filled pauses and hesitations) were strongly related to fluency ratings. Comprehensibility and fluency ratings were highly correlated, and fluency was related to comprehensibility more strongly than to accentedness.

Another stream of research included the factor of raters in the investigation of the correlation between utterance and perceived fluency. Kormos and Dénes (2004) explored variables predicting fluency perception of native and non-native speaking teachers and how they distinguished fluent L2 learners from non-fluent ones. Speech samples from 16 Hungarian L2 learners at two levels of proficiency were investigated with the help of computer analysis of pauses, in addition to measures of output quality such as accuracy and lexical diversity. Temporal and linguistic measures were correlated with the judged fluency scores by three native and three non-native teacher judges. It was concluded that SR, MLR, PTR and the number of stressed words per minute were the best predictors of fluency. However, the raters were not consistent in terms of how important accuracy, lexical diversity and the mean length of pauses were. The number of filled and unfilled pauses and other disfluency phenomena did not influence perceptions of fluency in this research. Rossiter (2009) explicitly aimed at comparing ratings of different groups of judges. Three groups of raters, namely novice and expert native speakers of English (undergraduate education students and experienced ESL teachers) and advanced non-native speakers were asked to judge the speaking fluency of 24 English L2 learners elicited from story narrating tasks at two time points (10 weeks apart). The ratings of the three rating groups were highly inter-correlated at both time points. Temporal measures of the number of pauses per second and pruned SR were correlated with perceived fluency. Pausing, self-repetition, SR and fillers accounted for three-quarters of the negative temporal impressions by listeners. Pinget et al.'s (2014) research investigated what native raters took into consideration when assessing L2 fluency. The relationship between temporal and acoustic features of speech and fluency ratings and accent was explored. The study found that acoustic measures (based on syllable length and pause length or frequency) showed high predictive value for fluency ratings. Two metrics of repair fluency (number of corrections and repetitions) showed a certain degree of predictive power but were non-negligible in contrast to findings of previous research (Cucchiarini et al., 2002; Kormos & Dénes, 2004). It should be noticed that measures of fluency in this research were measured based on spoken time instead of total time. SR and PTR were not included since they were regarded as confounded because they encompassed both speech speed and pausing features. Moreover, it was found that L2 residuals (partialling out L1 measures) were not better than traditional L2 measures for the prediction of perceived fluency, though they could predict a large proportion of the variance in fluency ratings.

It seems safe to say that previous studies found strong associations between utterance fluency and perceived fluency irrespective of differences in number and types of participants, types of speaking tasks and raters. Speed and pause measures were found to be good predictors of fluency ratings in most studies while repair measures seemed to play a limited role in or not influence perceived fluency (Bosker et al., 2012; Kormos & Dénes, 2004). Differences in previous studies also raise some methodological issues. De Jong et al. (2013) pointed out that studies relating listeners' perception to objective fluency measures run the risk of being circular since perceived fluency is dependent on instructions that the raters receive and notions that the listeners or raters have of fluency constructs. Another issue deserving attention is the possible confounding of some measures of utterance fluency, for example, the calculation of speech rate as words or syllables per total time (including pauses), with which method breakdown fluency and speed fluency might overlap with each other.

2.2 Fluency in interpreting

2.2.1 Cognitive factors in interpreting

Interpreting is a complex language processing task in which cognitive resources play an essential role. Cognitive abilities identified as important for language processing include, among others, lexical access and retrieval and memory abilities (Christoffels, De Groot, & Waldorp, 2003). Previous studies exploring cognitive factors affecting interpreting performance are summarised and reviewed below from three perspectives: lexical access and retrieval, working memory capacity and cognitive control.

2.2.1.1 Lexical access and retrieval and interpreting

Lexical access involves the recognition of linguistic forms and the arrival at the correct lexical entry from the mental lexicon and lexical retrieval involves the selection of lexical concepts that are subsequently transformed into linguistic forms (Levelt, 1989; Snellings, van Gelderen, & de Glopper, 2002). These two processes are essential for language comprehension and have been proved to be important for interpreting performance. Two translation routes are summarised by researchers: the transcoding route and the meaning-based route (e.g. Anderson, 1994; Bajo, Padilla, & Padilla, 2000; De Groot, 2000; Paradis, 1994). The transcoding route refers to the process of literal transposition of word or multi-word units whereas the meaning-based route involves the comprehension and conceptualisation of the input meaning which is the basis for production (Christoffels et al., 2003). The efficiency of the access and retrieval of words or meaning and the translation equivalents may be crucial for the task of interpreting, which involves both the comprehension and production phases.

It was found by Fabbro and Darò (1995) that students with SI training experience were more resistant to the disruption effect of delayed auditory feedback than participants with no SI background. In the same study, it was shown that student interpreters generally had higher verbal fluency than non-interpreters and did not show any significant speech disruption in either L1 or L2. The research of Bajo et al. (2000) compared interpreters, student interpreters, bilinguals and other professionals and found that interpreters had faster semantic access than other participant groups in a categorisation task. Christoffels et al. (2003) tapped into the correlation between lexical retrieval and memory abilities and SI performance of untrained bilinguals. Results showed that interpreting performance was correlated to the latency of lexical retrieval (elicited through a word translation task and a picture naming task) and associated with memory ability (tested through a reading span task in two languages and a verbal digit span task in the native language), though less strongly. However, in the exploration of Cai, Dong, Zhao, & Lin (2015) exploration of factors contributing to individual differences in the development of consecutive interpreting competence of beginner student interpreters, they failed to find a significant impact of lexical retrieval, elicited from a translation recognition task, on interpreting performance. Differences in lexical retrieval tasks and modes of interpreting might contribute to the divergence of these findings. The contribution of lexical access and retrieval in interpreting performance calls for further investigation.

2.2.1.2 Working memory capacity and interpreting

Highly demanding on working memory, interpreting is regarded as a process of maintaining equilibrium between different cognitive demands (Chernov, 2004). There have been some studies relating working memory capacity and overall interpreting performance. In the study of Timarová et al. (2015), working memory capacity was marginally significantly related to SI measures and only to such components with a predictable high memory component like figures and lists of nouns. However, working memory capacity was also found to support SI ability (Injoque-Ricle et al., 2015) and to be a strong predictor of SI performance (Macnamara & Conway, 2016). In terms of the relationship between working memory and fluency, there is quite ample evidence for a close relationship between cognitive abilities and L1 or L2 oral production (see section 2.1.1 for a review). Studies relating cognitive abilities directly to fluency in interpreting performance are scarce, but working memory was found to predict SI fluency in trainee interpreters' performance in the study conducted by Lin, Lv, & Liang (2018), which indicates the critical role of working memory capacity compared with language skills in fluent interpreting production. Previous studies on working memory of simultaneous interpreters, though yielding contradictory findings, shed light on the current research and are reviewed in the following section.

A considerable number of studies explored whether interpreters had an advantage regarding working memory. The hypothesis of the existence of interpreter memory advantage was verified by several studies (Christoffels, De Groot, & Kroll, 2006; Köpke & Nespoulous, 2006; Padilla, Bajo, & Macizo, 1995; Signorelli, Haarmann, & Obler, 2011; Tzou, Eslami, Chen, & Vaid, 2011; Wang, 2013) and rejected by several other studies (Chincotta & Underwood, 1998; Köpke & Nespoulous, 2006; Liu, Schallert, & Carroll, 2004). One of the earliest studies reporting an interpreter advantage in working memory was that of Padilla et al. (1995). Digit span, reading span and free recall with and without articulatory suppression tasks were used to test the memory of four groups of participants: ten professionals (half with 4-5 years of experience and the other half having just obtained their diploma), ten non-interpreter control subjects, ten beginning level student interpreters and ten advanced-level student interpreters. They found that the professional group outperformed the other groups in digit span, reading span and free recall with articulatory suppression, but not in free recall without articulatory suppression. There was no difference between the control group and the two student groups; a memory advantage of interpreters was thus observed, indicating that interpreters were less disturbed by phonological interference. An interpreter memory advantage was also confirmed by Christoffels et al.'s (2006) study, which explored basic language performance and working memory capacity, two aspects hypothesised to be important for SI. Three groups of participants were chosen: three professional interpreters, 39 unbalanced bilingual university students and 15 matched highly proficient English teachers. Word span, speaking span and reading span tasks were administered in order to test the memory of interpreters in both Dutch and English. They found that the interpreters outperformed the students and teachers in memory tasks, though not in language tasks. Köpke and Nespoulous (2006) compared the performance of 21 professional interpreters, 18 second-year novice interpreting students and two control groups in a series of tasks. Group effects were significant in the listening span task, the category probe task and the free recall with articulatory suppression task, but the novice interpreters outperformed the experts. Furthermore, there were no between-group differences in both the simple span task and the Stroop test. Signorelli et al. (2011) examined whether interpreters had an advantage in working memory and took into account different constructs of working memory and the possible influence of age. Participants in their study included 12 younger interpreters, 11 younger non-interpreters, 13 older interpreters and 11 older non-interpreters. The tasks undertaken involved non-word repetition, reading span and cued recall. Interpreters outperformed non-interpreters in reading span and non-word repetition, but not in cued recall and articulation rate. The results implied that interpreters manipulated information in working memory and processed sub-lexical phonological representations better than non-interpreters. Another study on interpreter memory advantage comes from the study conducted by Tzou et al. (2011). Digit and reading span tasks were administered to compare the memory performance of three groups of participants: 11 year-one student interpreters, nine year-two student interpreters and 16 bilingual controls. Student interpreters outperformed bilingual controls in reading span tasks in both L1 and L2. However, it should be noted that larger working memory spans were observed in participants with higher L2 proficiency compared with participants with lower proficiency, which indicates the contribution of L2 proficiency to interpreter working memory advantage. A more recent study investigating the working memory advantage of conference interpreters is that of Chmiel (2018), which is the first longitudinal study to examine the effects of training on the memory of trainee conference interpreters. Professional interpreters outperformed controls on all working memory span tasks (L1 and L2 reading span and L1 listening span) consistently and were not affected by stimuli modality and recall mode (serial vs. free). It was found that interpreter training, but not experience, improved working memory capacity and predicted interpreting performance. The meta-analysis of Mellinger and Hanson (2019) combined multiple studies on this issue and the results were indicative of a greater working memory capacity of professional interpreters than the comparison groups and the study observed an overall positive correlation between working memory capacity and measures of SI quality. The study of Wen and Dong (2019) also found evidence supporting an interpreter advantage in both working memory and short-term memory spans. Their meta-analysis found that such an advantage was more expressed in verbal memory tasks than in numerical, letter and spatial tasks, and interpreting expertise significantly moderated this advantage.

Despite the evidence from previous studies for an interpreter working memory advantage, no advantage for interpreters was found for digit span in the study conducted by Chincotta and Underwood (1998). Their study recruited 12 interpreting students (each with 100 hours of interpreting practice) and 12 bilingual English-major students.

Both groups were asked to recall lists of visually presented digits. The digit span task was conducted in both Finnish and English, and both with and without articulatory suppression. No group effect was found for digit span under any condition. Similarly, Köpke and Nespoulous (2006) observed no interpreter advantage for digit span or word span. Another study that failed to support an interpreter advantage in working memory was conducted by Liu et al. (2004). Three groups of participants were recruited: eleven professional interpreters, eleven advanced student interpreters (at the end of their second year of training) and eleven beginner student interpreters (at the end of their first year of training). The memory capacity of participants was measured with a listening span task. Significant group effects for SI performance were observed. However, the differences in working memory capacity among the three groups were not significant. The significant differences between the professional interpreters and the beginner and advanced trainee interpreters were attributed to the interpreters' task-specific skills and strategies, instead of general cognitive ability. Timarová et al. (2015) conducted an exploratory study testing the correlation between working memory capacity and professional SI measures. Working memory capacity was measured through letter span, Corsi task and complex span, and SI performance measures included local processing measures (lexical, syntactic and semantic processing) and global processing measures (vocabulary richness, ear-voice span and performance at different input rates). It was found that the relationship between working memory capacity and different SI measures were only marginally significant, and only for those with a high memory component. The study also found that age was negatively related to working memory capacity and general cognitive ability. The study of Wang (2016) study, with professional sign language interpreters as participants, failed to find significant correlations between bilingual working memory capacity and overall SI performance. The same trend was found for both language directions and for both native and non-native signers.

As reviewed above, we have not yet gained a clear and full picture of the role of working memory in interpreting. Exploration into whether interpreters have a memory advantage has not reached consistent findings. The contradictory results might be attributed to methodological factors (choice of tasks), procedural differences (different stimuli presentation duration and recall order) and inconsistent participation selection criteria (experience, age etc.) (Dong & Cai, 2015; Köpke & Nespoulous, 2006; Signorelli et al., 2011). Various memory tasks have been used, including simple span task, complex span task, free recall with or without articulatory suppression, non-word repetition and so on. It should be noted that the choice of different memory tasks might be an important reason for the discrepancies, since different tasks might have assessed different memory skills (Köpke & Signorelli, 2012). The number of participants in most of these studies has been relatively small, which is a limitation in terms of the consequently comparatively weak statistical power. Variables like L2 proficiency and age are found to affect interpreters' memory advantage (Signorelli et al., 2011; Timarová et al., 2015; Tzou et al., 2011) and should be taken into consideration in future studies. Moreover, the question of causality relationship between working memory capacity and SI quality has not been resolved (Mellinger & Hanson, 2019). Most studies have not provided reliable measures of the expertise of professional interpreters and have limitations regarding the control of subject variables (García, 2014). Generally, the correlation between working memory and SI is more common in the performance of untrained bilinguals and trainee interpreters than in professional interpreters. One explanation for this is that working memory is thought to be a predictor at comparatively lower levels of skill acquisition and plays an essential role when the skill is still not yet automatic (Timarová et al., 2015).

2.2.1.3 Cognitive control and interpreting

Cognitive control or executive control is "an umbrella term for the management of cognitive processes" and it includes "the ability to manage a complex set of task demands, to inhibit irrelevant or competing information, and to switch attention to goal-relevant information" (Dong, 2018, p. 190). Cognitive control mainly consists of three components of functions, i.e. inhibition, shifting and updating {Formatting Citation}. Shifting refers to the ability to shift between different tasks or mental sets (Sephen Monsell, 1996). Shifting ability is not only a "reflection of the ability to engage and disengage appropriate task sets" but also the ability to "perform a new operation in the face of proactive interference or negative priming" (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000, p. 56). Updating is related to replacing the old information that is no longer relevant to newer and more relevant information (Miyake et al., 2000; Morris & Jones, 1990). Inhibition is concerned with one's ability to "inhibit dominant, automatic, or prepotent responses when necessary" (Miyake et al., 2000, p. 57). The nature of simultaneity of comprehension and production of SI means that cognitive control may be more important than working memory capacity in this task. But empirical evidence for the direct relationship between cognitive control and interpreting is still isolated and limited.

Some evidence of a relationship between the central executive, an attentional control system of working memory, and SI performance has been provided. The study conducted by Timarová et al. (2014) administered four central executive tasks and three SI tasks to 28 professional interpreters. Four constructs of executive functions were tested, i.e. resistance to interference, resistance to automatic responses, updating and shifting. It was concluded that different functions of working memory predict different SI sub-processes in complex patterns. The inhibition function of the central executive of working memory seemed related to interpreting experience while other functions did not. The functions of shifting and updating did not show any association with interpreting experience. It was concluded that shifting and updating functions reflected cognitive abilities that were important for interpreting but did not seem to develop with practice.

Some studies were concerned with whether interpreting training would lead to an interpreter advantage in domain-general cognitive control, with most studies supporting the interpreter advantage in one or two functions of cognitive control. Some empirical studies have been conducted to lend support to the view that interpreting experience contributes to the shifting function of cognitive control. Yudes, Macizo, & Bajo (2011) found that professional interpreters did not perform differently from control bilinguals and monolinguals in the Simon task, which tested inhibition, but outperformed bilingual and monolinguals in the Wisconsin Card Sorting Test (WCST), which tested shifting. The study indicated that interpreting experience was associated with cognitive flexibility (mental set shifting), a finding which was echoed by Dong and Xie (2014). They compared four groups of young adult Chinese-English bilinguals who differed in L2 proficiency and interpreting experience. They found no significant group difference across all groups in the Flanker task which tested the inhibition function of cognitive control, but there was a significant difference between groups with varied interpreting experience in the shifting function. Their conclusion was that interpreting experience significantly contributed to mental set shifting enhancement among young adult bilinguals. Further evidence for the development of cognitive abilities during interpreting training comes from Dong and Liu's (2016) research. They compared three groups of young adult bilinguals who received, respectively, consecutive interpreting training, written translation training and general L2 training in a pre-test and post-test (one semester apart). Participants were tested twice on number Stroop task, switching colour-shape task and N-back task. Results indicated that interpreting training produced significant cognitive advantages in the efficiency of switching and updating while the translation training only led to marginally significant improvement in updating efficiency, which suggested that the experience of language switching under high processing demands brought more domain-general advantages. Becker, Schubert, Strobach, Gallinat, & Kühn (2016) provided evidence that simultaneous interpreters possessed cognitive advantages in cognitive control tasks. They investigated whether simultaneous interpreters had cognitive advantages in cognitive control tasks compared to a professional multilingual control group. The results revealed that the simultaneous interpreters exhibited less mixing cost in the task switching paradigm and showed a dual-task advantage.

Moreover, there are positive findings suggesting that interpreting experience is related to the enhancement of other functions of cognitive control. Morales, Padilla, Gómez-Ariza, & Bajo (2015) found that simultaneous interpreters outperformed a control group in updating skills measured through a dual version of the N-back task. In terms of attention networks tested through ANTI-V task, no group difference was found in conflict resolution, but the interaction between alertness and orienting networks differed between interpreters and non-interpreters. Research conducted by Hiltunen Pääkkönen, Vik, & Krause (2016) indicated that executive control seemed to play a significant role in the explanation of the cognitive processes of simultaneous and consecutive interpreting. They pointed out that the advantage in memory and executive control might be expertise-dependent and the differences between different language expert groups were thought to have been achieved as a result of thorough experience in their fields of expertise. Dong and Zhong (2017) found evidence for advantages in early attentional processing and monitoring using a Flanker task to explore how interpreting experience might modulate the executive functioning of young adults. The response time data showed a smaller interference effect for the group with more interpreting experience, indicating an advantage in inhibition. The overall results indicate that interpreting experience might enhance early attentional processing, conflict monitoring and interference suppression. The recent longitudinal study by Dong, Liu, & Cai (2018) tested N-back (non-verbal updating), L2 listening span and letter running span (verbal span) with two comparable groups of Chinese learners of English who received, respectively, consecutive interpreting or general L2 training for one semester. Results of the study showed that consecutive interpreting training enhanced updating efficiency while general L2 training did not. The efficiency of updating in both the pre-test and post-test predicted interpreting performance. The relationship between verbal spans and consecutive interpreting performance was weaker. The indicated interpreter advantage in updating was thought to be attributed to the shared underlying attention control process between updating and interpreting.

However, there have also been negative results of explorations into the relationship between general-domain cognitive control and interpreting performance. Research by Babcock and Vallesi (2017) showed that interpreters did not continue to garner benefits from bilingualism and seemed to possess benefits specific to SI experience. Through memory tests, the colour-word Stroop task, the attention network test and a non-linguistic task-switching paradigm, they examined professional interpreters and matched bilinguals. It was found that interpreters did not show advantages in conflict resolution or switching cost, where bilingual advantages were noted, but an interpretation-specific advantage was found regarding mixing cost in the task-switching paradigm and interpreters had a verbal and spatial memory advantage. Rosiers, Woumans, Duyck, & Eyckmans (2019) compared the working memory capacity and executive functions of student interpreters with two other groups of advanced language users who were all at the beginning of Master training. Results showed only negligible differences between the three groups at the onset of training. The study failed to find the presumed cognitive advantage of aspiring interpreters in terms of executive control.

As can be seen above, studies generally support that cognitive control is important for SI performance, mainly in terms of the functions of shifting (Timarová et al., 2014) and updating (Dong et al., 2018; Timarová et al., 2014). The evidence overall suggests that interpreting expertise enhances semantic processing, working memory and cognitive flexibility (García, 2014). Interpreting training or experience is supposed to contribute to the enhancement of the functions of shifting (Becker et al., 2016; Dong & Liu, 2016; Dong & Xie, 2014; Yudes et al., 2011), updating (Dong & Liu, 2016; Morales et al., 2015) and inhibition (Dong & Zhong, 2017). Though some empirical evidence has been gathered supporting the notion of interpreter advantage in one or more functions of cognitive control, some studies have failed to find evidence supporting the relationship between cognitive control and SI performance or interpreter advantage (Babcock & Vallesi, 2017; Rosiers et al., 2019). Research findings regarding the relationship between cognitive control and interpreting have been inconclusive. It has been pointed out that interpreter advantage in terms of cognitive control might be the effect of interpreting training and experience (Hiltunen et al., 2016; Rosiers et al., 2019). More empirical studies are needed to substantiate the current findings, to explore the contributing factors of interpreters' cognitive advantage, and to confirm the role of different cognitive functions in interpreting performance. Longitudinal studies on the roles of cognitive abilities in the development of interpreting expertise are still rare and may be conducted in the future.

2.2.2 Utterance fluency in interpreting

Previous studies exploring fluency in interpreting have given much weight to utterance fluency in interpreting output and have mostly investigated the dimensions of breakdown and repair fluency. Related studies on disfluencies and pauses in interpreting are reviewed in this section.

2.2.2.1 Disfluencies in interpreting

Gósy (2007, p. 93) defined disfluency as "the phenomena that interrupt the flow of speech and do not add propositional content to an utterance". Speech disfluencies reflect the increase in cognitive effort demanded by lexical or syntactic uncertainty, planning, or production problems (Shreve, Lacruz, & Angelone, 2011). Disfluency in interpreting mirrors difficulties and uncertainties during the cognitive processes of interpreting. The study of fluency in interpreting could inform professional development and training pedagogies because research findings could reflect difficulties facing interpreters. Analysis of these speech phenomena may offer a unique opportunity to understand the spoken translation product as a direct reflection of underlying cognitive processes (Shreve et al., 2011). The taxonomy of disfluency is not consistent in previous studies (Bakti, 2009; Bendazzoli, Sandrelli, & Russo, 2011; Cecot, 2001; Tissi, 2000). Main types of disfluency identified in past research include repetitions, repairs, filled pauses and false starts. Previous studies on fluency in interpreting have explored different types of disfluencies through both qualitative and quantitative analyses and have been concerned with their distribution in interpreting output, the comparison of disfluencies in source speech and target interpreting output, and variables that may affect fluency in interpreting output, such as input rate and accent.

Pöchhacker (1997) conducted a case analysis based on a live broadcasting SI of the US president's speech in Germany and explored aspects of accent, voice, fluency of delivery, cohesion and consistency, completeness and correctness, providing a good example of quantitative analysis of fluency. This study described features of fluency from different aspects, including the duration and frequency of pauses longer than 2 seconds, tempo (speech rate), voiced hesitations and false starts. It was found that most pauses in the interpreters' output reflected the time lag at the beginning of utterances and only one-third of interpreters' pauses could be clearly identified as hesitations. The study stated that live broadcast media interpreting was possible at a level with few inconveniences to the audience, though the requirements for this kind of SI were particularly stringent. Pöchhacker (1995) also examined the speeches at a three-day conference and the SI output from the perspective of slips and shifts. The categories of corrected and uncorrected slips and structure shifts (false starts, lexical blends and syntactic blends) were included. It was found that there were more slips and shifts in the interpreters' output compared to those in the speakers' output. False starts constituted a high proportion of these, irrespective of speaker or language direction, which were seen as a universal of speech production instead of a unique characteristic of SI. An SI-specific functional taxonomy of disfluencies was proposed by Tissi (2000) in her descriptive analysis of silent pauses and disfluencies in SI, with ten student interpreters as participants. Quantitative analyses were conducted to explore possible correlations between the occurrences of pauses and types of disfluencies in the source and target texts. She worked with two major categories to describe non-fluencies: silent pauses (grammatical and communicative pauses and non-grammatical pauses) and disfluencies. The latter included filled pauses (vocalised hesitations and vowel and consonant lengthenings) and interruptions (repeats, restructuring and false starts). It was found that vowel and consonant lengthenings were much more abundant in the target texts, and false starts occurred only in the target texts. The general findings were against the presence of a systematic source-target text correspondence of non-fluencies. The communicative and strategic use of some of the non-fluencies by the interpreters was stressed. The variable of input rate was included in Cecot's (2001) research, which analysed occurrences of non-fluency in the SI performance of eleven professional interpreters who translated two texts at two different input rates from English to Italian. The two major categories of non-fluency analysed were disfluencies (filled pauses, repeats, restructuring, false starts, vowel and consonant lengthenings) and unfilled pauses (segmentation, rhetorical and hesitant pauses). Non-fluency occurrences in source and target texts were compared and it was found that interpreters tended to follow the speaker's pattern of unfilled pause occurrences. In terms of distribution, it was found that disfluencies outnumbered unfilled pauses in both texts of both rates. The study highlighted that there was a need to raise the awareness of interpreters' delivery since significant inconsistencies existed between objective performance and interpreters' subjective perception. Concerned about what different types of disfluency signalled from the perspective of speech production, Bakti (2009) analysed the distribution of speech disfluencies in the examined corpus of trainee and professional simultaneous interpreters working from English (B language) into Hungarian (A language). Her study differentiated two groups of disfluencies, i.e. disfluencies caused by uncertainty (hesitations, fillers, repetitions, restarts, lengthenings and pauses within words) and error-type disfluencies (Freudian slips, grammatical errors, contamination, false word activation, slips of the tongue etc.). In her research, restarts and grammatical errors were the most frequently occurring error-type disfluencies, followed by grammatical errors and false word activation. It was pointed out that restarts signalled coordination problems between lexical access and articulatory planning, and grammatical errors indicated problems in morphological and syntactical planning and might be related to the availability of the processing capacity for the task (Gile, 2009). Fu (2013) explored the effects of interpreting directionality on disfluencies, i.e. ungrammatical pauses (within-phrase pauses with durations of no shorter than 0.25s), repeats and repairs, in the consecutive interpreting output. Results of quantitative analyses showed a significant impact of directionality on disfluencies in interpreting output, but the correlations between directionality and fluency were inconsistent and language-specific. There were more occurrences and a higher ratio of inappropriate pauses in A-B output than in B-A output, while repairs witnessed the opposite trend.

The greater number of repairs in B-A output was attributed to the advantage of the mother tongue, which might have triggered excessive repairs. The study did not offer unanimous results to support the hypothesis that the consecutive interpreting output in the mother tongue was more fluent than that in the foreign language.

Some studies focused on a specific type of disfluency in the interpreting output, for instance repairs. Petite's (2005) research studied repairs and self-monitoring in SI. The data source was the performance of eight professional conference interpreters, working from B to A language and recorded at four different international conferences on topics of general interest. Based on Levelt's (1983) classification of repairs, the analysis included three types of post-articulatory repairs, i.e. A-repairs (A for appropriateness), E-repairs (E for error) and D-repairs (D for different order of words), and an additional category of mid-articulatory repairs. Interpreters spent processing resources on the production of repairs including non-obligatory ones and they tended to repair even when the repair might not be cost effective. The research concluded that repairs were not necessarily the results of the detection of an error and interpreters made repairs according to their own standard of acceptability. Quantitative analyses revealed that the majority of repairs were output-generated, and the total number of E-repairs was lower than that of A-repairs, D-repairs and mid-articulatory repairs. This research provided a good example of using an authentic corpus to tap into repairs, one of the main types of disfluency. Using data retrieved from two national interpreting contests in China, Zeng and Hong (2012) outlined a modified system of self-repairs in interpreting based on previous frameworks (Kormos, 2000; Levelt, 1983) and explored patterns and features of repairs of trainee interpreters in Chinese-English consecutive interpreting. His results showed that the repair rate of content-related errors was higher than that of form-related errors, assumingly due to insufficient competence of trainee interpreters. Modes of repairs were subject to the influence of task types and demands, indicated by more repairs for appropriateness in dialogue interpreting and more repairs of content in conference interpreting.

2.2.2.2 Silent pauses in interpreting

Pause, an important sub-parameter of fluency, is another focus of the investigation into fluency in interpreters' output. Silent or unfilled pauses in interpreting reflect "highly directed, sometimes exclusive attention to input" (Setton, 1999, p. 246). The judgment of whether a silent pause is disfluent is subjective and the standard may not be consistent among different audiences or on different occasions. This is probably the reason why silent pauses were not included in the category of disfluency in relevant studies (Cecot, 2001; Tissi, 2000). Previous studies on pauses in interpreting involve the exploration of the frequency, duration, syntactic distribution and range of pauses.

Research by Tissi (2000) on silent pauses and disfluencies in SI explored pauses in both source and target texts of the SI output of student interpreters. In her research, filled pauses (vocalised hesitations and lengthenings of vowels and consonants) were categorised under disfluencies, and silent pauses were divided into grammatical and communicative pauses and non-grammatical pauses. Quantitative analyses led to the conclusion that the same number of pauses per minute was produced regardless of the number of pauses in the source speech. The interpreting output contained fewer but longer silent pauses than the source speech, and target texts had a slightly higher number of grammatical pauses. The reduced frequency of pauses in SI was attributed to the fact that the source speech was spontaneous. In terms of the range of pauses, there was a much higher frequency of pauses within the range from 0.25 to 1.25 seconds for both the source and target texts, but target texts showed much higher incidence of pauses within the range of 2.5 to 5 seconds and 5 seconds up, which had no counterparts in source texts. Cecot's (2001) study on non-fluency in professional interpreters' SI performance found a decrease in hesitation pauses and an average increase in the number of filled pauses, corrections and lengthenings of vowels and consonants when the source speech rate was higher. The most frequently used pauses were segmentation pauses, followed by hesitation pauses. In Ahrens's (2005) research on prosodic features in SI from English to German in an authentic corpus of professional simultaneous interpretation, target texts were found to show fewer but longer pauses than those in the source texts. The author attributed the long pauses in the target texts to the waiting for more information and higher cognitive load. The research results contradicted the finding of Alexieva (1988) which found fewer and shorter pauses in the output of student interpreters. The research conducted by Wang and Li (2015) explored characteristics and motivations of pauses in Chinese-English SI output, comparing the performance of five trainees and five experienced interpreters. They found that compared with the original speech, the SI output contained less frequent but longer pauses. The distribution of pauses was hierarchical and corresponded to syntactic complexity except that the frequency of pauses inside phrases was markedly high. Half of the pauses in each group occurred before sentences and before clauses, which confirmed earlier findings (Cecot, 2001). About a quarter of pauses in both groups occurred inside phrases, which was deemed unnatural in spontaneous speech. Expert interpreters produced fewer and shorter pauses than trainees and their pauses were more appropriate and mostly occurred at major syntactic junctions, which was regarded as an indication of expertise development and was possibly due to the better segmentation skills and longer EVS of the expert interpreters. It was pointed out that unnatural pauses in C-E SI output revealed both the encountered difficulties and the decision-making mechanism of interpreters.

Studies on pauses in consecutive interpreting also abound. Mead (2000) examined pauses of 15 final year trainee interpreters in their consecutive interpreting performance into their A (Italian) and B (English) languages. Performance in the two language directions was compared, and indexes of pause duration (total pause duration and average pausing time per minute) and phonation time ratio in the target interpreting production were measured. Statistical analyses found that the fluency of trainee interpreters in their A and B language interpreting output differed significantly. Filled pauses and total pauses (filled and silent pauses) were significantly higher in B language than in A language output, and there was a significant negative correlation between silent and filled pauses in English. Mead (2002) also compared the consecutive interpreting performance of interpreters of different levels of experience and in different language directions (A-B and B-A). Groups of participants included beginner and advanced student interpreters and professional interpreters. The study found that professional interpreters had fewer problems with information formulation but made more hesitations with no apparent reason, which was supposed to reflect better extralinguistic skills and strategies and the automatic processes. In Mead's (2000, 2002) research into underlying reasons for pauses and hesitation in consecutive interpreting, retrospective explanations for pauses of participants were investigated and sorted into five categories (formulation difficulties, difficulties with notes, logical doubts, no apparent perceived reasons and others). The studies indicated that linguistic knowledge (like lexical selection), extralinguistic competence (e.g. analytical listening, note-taking) and interpreters' use of strategies to manage difficulties all contributed to fluency. Xu (2010) conducted an empirical study on pauses in the Chinese-English consecutive interpreting performance of five professional interpreters and explored the duration and frequency of different types of pauses in the interpreting output. The study summarised twelve types of pauses according to four categories of causes (information, production, strategy and others) elicited from stimulated recall. The research found that pauses in consecutive interpreting output were closely related to online cognitive processing and interpreters' adoption of strategies. Notes searching led to pauses of the highest frequency and message retrieval in the target language caused pauses with the longest duration. Using language directionality as the variable, Fu's (2012) study tapped into features of pauses and correlation between fluency and directionality in the consecutive interpreting performance of student trainee interpreters, and compared the performance of student interpreters with the performance of the two recorded professional interpreters. The study found that pauses were important during the online cognitive processing of interpreters, playing multiple roles including buffering cognitive load, marking syntactic boundaries, deriving rhetorical effects and so on. Results of the study did not corroborate the hypothesis that interpreting output in the mother tongue was more fluent than that in the foreign language, although directionality did affect the pausing features in the output. It was also found that pauses tended to be reflections of disfluencies for trainee interpreters and of the successful use of strategies for professional interpreters, but the amount and frequency of pauses did not differ significantly between trainee and professional interpreters.

As discussed above, most previous studies on utterance fluency in interpreting are descriptive and mainly investigate disfluencies and pauses in interpreting output, providing valuable references for the current exploration of fluency in SI. Generalisation from these studies is to some extent difficult because of the variation in the taxonomy of disfluencies, the selection of influencing variables, the methods of analysis, the modes of interpreting and so on. In terms of the taxonomy, it should be noted that filled pauses are regarded as one category of disfluencies while silent pauses are not. The reason is that filled pauses are supposed to have negative effects on the evaluation of fluency in interpreting (Pradas Macías, 2006) while silent pauses have multiple functions in language production, which can be grammatical, communicative or hesitative. More variables that might influence a fluent interpreting production should be included in future research. Research on the effects of directionality and interpreting expertise on utterance fluency in interpreting is still scarce (Fu, 2013; Wang & Li, 2015) and more empirical studies are required to substantiate the current findings. The causes and functions of disfluencies and silent pauses in interpreting output, for instance the communicative and strategic use of them in the process of interpreting, should be further explored.

2.2.2.3 Perceived fluency in interpreting

Quality is seen as a function of communication efficiency, and "both package and content should be optimized" so as to fulfil the objectives of the message sender (Gile, 1991, p. 199). Interpreting quality may be assessed from various subjective perspectives (Pöchhacker, 2001). The importance of fluency as one of the essential quality criteria for interpreting assessment has been recognised by a number of surveys in which users of interpreting services were asked to comment on which aspects of interpreting output were most valued by them (Bühler, 1986; Kurz, 1993, 1994, 1996; Moser, 1996). Fluency is ranked by different user groups as important consistently, after accuracy and fidelity (Han, 2015a).

When exploring the quality of interpreting, the evaluation, measurement and assessment of interpreting quality should be distinguished according to different purposes (Moser-Mercer, 1996). Evaluation is more about the judgment of the interpreting services in a natural setting, measurement is mostly concerned with the analysis of the interpreting output in a laboratory setting, and assessment is mainly for interpreting training or certification of the interpreting profession, though sometimes evaluation and assessment are combined. Perceived fluency explored in the current research refers to the assessment and perception of utterance fluency in interpreting output. Quality perception of interpreting is not synonymous with quality assessment and is mainly based on user experience (Cheung, 2013). It is pointed out that the quality of interpreting services should be judged from the perspective of listeners and evaluated according to their expectations (Kurz, 2001; Moser-Mercer, 1996; Seleskovitch, 1986). Variables that may have effects on the end user include speed, pauses, hesitations, repairs, intonation, register, cohesion and so on (Kurz, 2001). The perception of interpreting quality is subjective, but the analysis of interpreting product should be complemented with the subjective judgment of listeners (Pöchhacker, 1994).

Previous studies on user perception of interpreting have gathered extensive observational data, though with little comparability between different studies (Kurz, 2001). Surveys and questionnaires are the most used research tools for user expectations and evaluations. Fluency was placed fifth out of eight items in the overall ranking ahead of correct grammatical usage, native accent and a pleasant voice in the classic survey conducted by Kurz (1993) on how different user groups and interpreters rated various features of conference interpreting. The study adopted the first eight criteria of user perception introduced by Bühler (1986), i.e. completeness of interpretation, correct grammatical usage, fluency of delivery, logical cohesion of utterance, native accent, pleasant voice, sense consistency with original message, and use of correct terminology. The series of studies of Kurz (1993, 1994, 1996) found that conference interpreters and different user groups differed in the assessment of some criteria (correct grammar, pleasant voice, native accent), though there was quite high agreement among all groups on some other criteria (sense consistency, logic cohesion, correct terminology). The background of users of interpreting service was taken into account by Kopczynski (1994) in a survey of their attitudes and expectations of interpreting services, and speakers and listeners were distinguished. Speakers regarded fluency as the most important factor while style and fluency were the most important for listeners, though both speakers and listeners regarded incorrect terminology as the greatest irritant. The survey conducted by Moser (1996) and funded by AIIC confirmed that different user groups had different expectations of interpretation. Conducted at 84 meetings, the survey covered both speakers and listeners at different types of conferences. Factors of user experience and gender of users were considered. Fillers were found more irritating by women than men and monotonous delivery was found irritating (cited in Kurz, 2001). In a recent survey involving 704 AIIC interpreters on the quality of conference interpreting conducted by Pöchhacker (2012), fluency was regarded as very important by 71% of the participants, and ranked third out of eleven quality criteria, behind sense consistency and logical cohesion. While most of the studies in this line were about SI, the users' expectations of and responses to consecutive interpreting performance were explored in the case study conducted by Marrone (1993). Several parameters of interpreting quality were investigated using a questionnaire, including information completeness, style quality and correct terminological usage, and quality of intonation and delivery. Besides, some miscellaneous issues were analysed, e.g. delivery speed, unpleasant delivery, reproducing speaker's faults.

Some studies focused on certain variables related to fluency in the interpreting output in controlled experiments. Following previous research, Pradas Macías (2006) provided some empirical evidence of the role of salient silent pauses in expert users' evaluation of the criterion of fluency in SI. Exploring the quality expectations and evaluation of 43 expert users of German-Spanish SI performance manipulated by additional silent pauses (2-6 seconds in duration), the study suggested that unfilled pauses, as a sub-parameter of fluency, were supposed to have negative effects on the evaluation of SI fluency, though the results were not statistically significant. Pradas Macías argued that silent pauses in SI output were related to different parameters, such as intonation and speech continuity, and concluded that the way the study isolated and tested sub-parameters would promote a deeper understanding of SI quality and its perception.

Another line of research directly relates measures of interpreters' utterance fluency to measures of users' perception of fluency in interpreting output. This line of research offers insights into what features of an interpretation actually constitute and affect judged fluency, providing help to trainee interpreters to cope with difficulties during the learning process and potentially contributing to the development of automatic assessment of interpreting (Yu & van Heuven, 2017). The correlation between (para)linguistic parameters and perceived fluency by human raters in English-Chinese (B-A) SI performance was investigated by Han (2015a). Using a group of 32 practicing interpreters as participants, his study explored three underlying dimensions of utterance fluency which involved nine dimensions of fluency, including speed fluency (speech rate, phonation time ratio and mean length of run), breakdown fluency (number of unfilled pauses, mean length of unfilled pauses and total number of pauses) and repair fluency (number of false starts, reformulations, and replacements). Fluency ratings were positively correlated with speed fluency measures (speech rate, phonation time ratio, and mean length of run) and negatively correlated with breakdown fluency measures. The best possible predictor of interpreters' fluency was phonation time ratio. Despite limitations regarding the number of coders and sample size, the study revealed that parameters of speed fluency were more related to perceived fluency. The correlations between judged fluency and quantifiable measures of utterance fluency and between judged fluency and accuracy in consecutive interpreting were explored by Yu and van Heuven (2017). Two consecutive interpretations of the same source speech by 12 trainee interpreters from Chinese (A language) to English (B language) were judged by ten raters. Twelve acoustic measures of utterance fluency were measured. The study found a close correlation between ratings for information accuracy and grammatical correctness and the judgment of fluency metrics (delivery speed, pause control, disfluency control and overall fluency). There were also significant correlations between the judged fluency criteria with most of the acoustic fluency measures. Results of linear regression analyses showed that the four criteria of judged fluency in consecutive interpreting were best predicted by effective speech rate. The author attributed this to the fact that all three aspects of fluency, i.e. speed, breakdown and repair fluency, were involved in effective speech rate. Fluency ratings were also related to the number of filled pauses, articulation rate and mean length of pauses. The study had implications for automatic and quantitative assessment of fluency in interpreting and interpreting teaching practices.

As shown from the research discussed above, it is generally acknowledged in past research on interpreting perception that fluency is regarded as an important constituent of interpreting quality by different user groups. Studies on the assessment of interpreting have been conducted from diverse perspectives: the product and performance of interpreting (Hatim & Mason, 1997), the process and product-oriented assessment (Gile, 2001), assessment for interpreter training purposes which includes aptitude testing for selection, intermediate formative testing, summative testing for degree purposes or certificate conferral, and ipsative assessment (Sawyer, 2004). However, there has been a lack of systematic methods and consistent practices in the assessment of interpreting, reflected by arbitrary selection of content, inconsistent administration of tests and failure to establish objective scoring criteria (Liu, 2015b; Liu, Zhang, & Wu, 2008; Sawyer, 2004). Constructs of utterance fluency are found to be correlated with perceived fluency in interpreting, but relevant studies are still few and suffer from low comparability. Future studies may try to establish consistent and objective scoring criteria when assessing fluency and explore more variables that might affect the perception of fluency in interpreting. More research on the relationship between utterance and perceived fluency in interpreting is also needed to corroborate the current findings.

2.3 Factors influencing fluency in interpreting

Factors that affect interpreting performance may involve, among others, general linguistic and extralinguistic knowledge and availability, cognitive abilities, the use of coping tactics or strategies, features of source speech (input rate, accent etc.), interpreters' experience, psychology and so on (Gile, 2009). Two aspects important for fluency in trainee interpreters' SI output were manipulated in the experiment conducted in the current research, i.e. training and input rate. This section reviews relevant studies on the influence of these two factors on interpreting and how it is related to the current research.

2.3.1 Training and interpreting expertise development

Formal training in Translation (interpreting and translation) has two important functions: to help trainees to "enhance their performance to the full realisation of their potential" and to help them "develop their Translation skills more rapidly than through field experience and self-instruction", apart from the social or professional functions of training programs (Gile, 2009, p. 7). Training is designed to develop expertise as a complement to linguistic ability and brainpower, which implies "special knowledge combined with highly effective or efficient procedures for performing a complex task, within certain parameters" (Setton & Dawrant, 2016, p. 314).

Scholars are interested in the differences in interpreting performance between experts and novices as a way to have a clear picture of the process of expertise development. The interpreting domain shares many common features with other domains of professional expertise (Ericsson, 2000). Differences between experts and novices range from "chunking of information, to reasoning, to speed of processing, to individuals' knowledge base and its organization" (Moser-Mercer, Frauenfelder, Casado, & Künzli, 2000, p. 108). Bransford, Brown, & Cocking (2000) summarised six key principles of expert knowledge: more features and meaningful patterns of information noticed, a great deal of organised knowledge acquired which reflects a deep understanding of the subject matter, experts' knowledge reflecting contexts of applicability and conditionalised to a set of circumstances, flexible retrieval of important aspects of knowledge with little attentional effort, a thorough understanding of the discipline, different levels of flexibility in response to new situations.

Two aspects of expertise factors that are key to expertise were identified by Ericsson and Smith (1991), i.e. acquired mediating mechanisms and deliberate practice. The mediating mechanisms in interpreting are knowledge schemas or cognitive structures and the ability to manage them (Ericsson, 2000). It was shown that experts had better-organised knowledge and they made greater use of recognition and retrieval than novices (Ericsson, 2014; Ericsson & Kintsch, 1995). It was stated that expert interpreting is mediated "not by fully automatic translation processes but rather by mechanisms and mental representations that provide interpreters with tools to gain more control over their translations" (Ericsson, 2000, p. 204). SI is only possible with the support of side channels including context, previous knowledge, cues from the environment and paralinguistic or pragmatic cues in speech, which allows the interpreter to synthesise meaning and get enough information to anticipate the general thrust and furnish the product with cohesive material. The interpreter needs to have "the trained mental procedures to make these syntheses, adequate mental schemas to (re)cognise the incoming concepts, and the ready language to produce these syntheses"

(Setton & Dawrant, 2016, p. 315). Novices in translation and interpreting are engaged in tactical learning of specific rules for solving problems; the tactical knowledge then becomes better organised and novices develop a set of strategies to solve the problems they encounter; as novices become more expert, their abilities to store problem information in long-term memory and to retrieve it are enhanced (Moser-Mercer et al., 2000).

Anderson (2015) describes three stages of skill acquisition, i.e. the cognitive stage, the associative stage and the autonomous stage. In the cognitive stage, novice learners develop the skill of declarative encoding; during the associative stage, errors in initial understanding are detected and eliminated gradually and element connections required for successful performance are strengthened; in the autonomous stage, the procedures become more and more automated and rapid. Moser-Mercer et al. (2000) hypothesised that shifting from consciously controlled to more automatic processing leads to a e in processing efficiency and speed, or that this shift results in a restructuring of the process itself. The initial gains in interpreting skill may be attributable to the increased automation of lower-level processes and the development of expertise at a higher level requires the restructuring of higher-level processing (McLaughlin, 1995; Moser-Mercer et al., 2000). In language learning, restructuring provides an explanation for cases of U-shape developmental functions where performance declines with more complex internal representations replacing less complex ones and increases again as skills become expertise (McLaughlin, 1995). This might also explain examples of trainee interpreters with sufficient aptitude starting to plateau without reaching the required level in the prescribed time (Moser-Mercer et al., 2000).

The other key factor for the development of interpreting expertise is deliberate practice. Personal practice apart from classroom working sessions plays a vital part in interpreter training. Classroom working sessions are essentially tutorials with the aim to guide students to acquire the appropriate strategies and develop the relevant skills (Gile, 2005). Deliberate, concentrated and task-relevant practice seems to be an essential part of the acquirement and maintenance of expertise (Ericsson & Smith, 1991). Deliberate practice refers to "highly targeted forms of individual training that focus on weak points, involving repetition and coaching", and is more taxing than casual practice without particular attention to the choice of task and materials (Setton & Dawrant, 2016, p. 47). Conditions of effective practice include tasks being well-defined, practice being at an appropriate level of difficulty for the individuals at their current stage of development, focusing more on individuals' difficulties and weaknesses, a social context of informative feedback, sufficient opportunities for repetition and correction of errors and so on (Ericsson & Lehmann, 1996; Shadrick & Lussier, 2009). Several studies have found a consistent correlation between attained level of performance and amount and quality of deliberate practice in a wide range of domains (Ericsson, 2001, 2002, 2014). The superior performance of experts and its complex mechanisms have been found to be the result of gradual improvements through deliberate practice over many years (Ericsson, 2000; Ericsson, Krampe, & Tesch-Romer, 1993). These deliberate efforts involve problem-solving and better methods of task performance. A prerequisite of deliberate practice is to engage in an activity with the primary goal of improving some aspect of performance (Ericsson, 2000). The acquirement of interpreting skill requires ample practice in order to produce the cognitive changes facilitating the circumvention of cognitive constraints and to complete the transition from novice to expert (Clark, 2008; Moser-Mercer, 2008). Before and during the course of attaining the level required for conference interpreting, active and passive language proficiency and general knowledge will need constant and regular work (Setton & Dawrant, 2016).

Training has important functions in the development of interpreting expertise and the enhancement of interpreting performance. By investigating the process of expertise development, describing the differences between expert and novice performance and pointing out key factors for expertise development, i.e. mediating mechanisms and deliberate practice, past research has provided important theoretical bases for exploration into the development of interpreting expertise and the effects of interpreter training on interpreting performance. Previous studies have paid attention to the development of cognitive abilities in interpreting training and correlations have been found between cognitive factors and interpreting performance (see a review in section 2.2.1). The current research pays attention not only to fluency performance per se, but also to the development of fluency in interpreting. Training, as an important factor in the development of interpreting performance, is included in the experimental design of this research.

2.3.2 Input rate as a variable of fluent interpreting output

Although input rate may not constitute an absolute comprehension problem, it is "a major constraining factor during SI", but the effects of input rate on SI have yet to be substantiated empirically (Seeber, 2011, p. 186). Previous research has generated divergent findings on the impact of input rate. A systematic study on this issue would help to enhance the understanding of the cognitive processes of SI, to choose appropriate SI training materials to suit trainee interpreters at various developmental stages and to manipulate difficulty levels of input materials to meet the demands of interpreter performance assessment (Han & Riazi, 2017; Liu & Chiu, 2009). The following section reviews previous studies on the impact of input rate on SI (on fluency in particular) since input rate is an important variable in the design of this research.

It has generally been believed that a high input rate had adverse effects on interpreting quality. Input rate of the source text in the range of 100 to 120 words per minute was regarded as ideal for interpreters (Pöchhacker, 2004). Findings of previous research provided support to the view that high input rate was a constraining factor for SI (Gerver, 1969/2002; Meuleman & Van Besien, 2009; Pio, 2003). Gerver (1969/2002) found that faster input speed led to longer lag and caused more errors and omissions and higher pause-to-speech ratio in interpreting output in his study involving six professional interpreters interpreting from French into English at different input speeds. It was also found that interpreters tended to increase their EVS (ear-voice span) without changing the output rate when the source speech was delivered at a higher rate. Pio (2003) considered two perspectives of SI quality, i.e. meaning equivalence between the source and target texts and fluency of interpreters' output, in her exploration of the relationship between source text delivery rate and SI quality with ten students and five professional interpreters as participants. Three categories were adopted to examine the fluency of the interpreted texts: phonation errors; unfilled pauses and filled pauses; repetitions, corrections and false starts. The results showed that there were more filled pauses, unfilled pauses, corrections and phonation errors under high input rate and the number of repetitions was slightly higher in the interpreted output for the slow source text. The interpreting for the fast source text was generally less fluent than that for the slow one. Different categories of fluency in the interpreting output of student and professional interpreters were compared, indicating that interpreters with different levels of expertise were affected differently by the input rate. In their study of how professional interpreters coped with high delivery speed when interpreting from A to B language, Meuleman and Van Besien (2009) also found negative impacts of high input rate on SI performance. More recent evidence is provided by the corpus-based research of Plevoets and Defrancq (2016), which demonstrated the effects of lexical density and the delivery rate of source speech on disfluencies in interpreting. Their research measured the information load in interpreting through modelling the occurrence of the filler, "uh(m)". Four measures of the information load (delivery rate, lexical density, numeral percentage and average sentence length) were explored by analysing and comparing a corpus of interpreted (target) and non-interpreted texts. Comparisons were made between non-interpreted and interpreted texts and between target and source texts. The input rate of source texts was shown to be the main predictor of "uh(m)s" in the target texts.

However, the issue is still inclusive. Some studies have reached mixed or different conclusions, i.e. that interpreters perform better under conditions of high input rate (Han & Riazi, 2017; Shlesinger, 2003). Shlesinger's (2003) study started with a counterintuitive hypothesis that interpreters performed better at a higher presentation rate. Sixteen professional interpreters interpreted the same six source texts twice in two sessions from B to A language, each text at two different delivery speeds. The texts were manipulated to contain four attributive modifiers before a noun head. Results showed that performance (modifiers rendered correctly) under conditions of high input rate was consistently, though not significantly, better than performance at the slower rate. It was supposed that a possible explanation was that higher presentation rate allows less time for the unrehearsed message to decay. It was pointed out that certain norm-driven strategies seemed to play a greater role than suspected. Liu et al. (2004) found that the SI performance for only one source speech was significantly worse for the fast delivery rate than for the slow one, and in general the delivery speed of source speeches did not affect participants differently. Chang's (2005) study yielded mixed results, her findings showing that high input rate reduced the accuracy of propositions significantly, but not the quality of the target language. In his research exploring directionality in SI, professional interpreters interpreted from English to Chinese and vice versa. The performance of the interpreters with Chinese as A language was significantly different at different delivery speeds for both the Chinese and the English speeches, while for the English dominant group no significant difference was observed at different speeds. Han and Riazi (2017) investigated the effects of speech rate and accent on SI using a mixed methods approach with a larger sample size of 32 professional interpreters. It investigated the impact of fast delivery speech on three measures of SI performance: information completeness, fluency of delivery and target language quality. It was found that the interpreting output was more fluent under conditions of fast delivery speed according to raters' assessment. Further analysis of the five fluency parameters revealed that the fast input rate tasks witnessed fewer pauses, higher speech rate, articulation rate and PTR (phonation time ratio), and longer MLR (mean length of run), which differed from those under low input rate conditions significantly.

As reviewed above, previous studies on the impact of input rate on SI have not reached consistent conclusions, so more empirical studies are required to substantiate these findings. Previous studies on the effects of input rate on interpreting performance have focused on different aspects of SI performance, including dimensions of information completeness and accuracy, fluency and so on, with a much greater weight on fidelity. For those studies which explored the effects of input rate on fluency in SI (Han & Riazi, 2017; Pio, 2003; Plevoets & Defrancq, 2016), the inconclusiveness of the findings might be attributed to the diversity of the explored constructs of fluency, different input rates chosen and the small sample size in most studies, which was not large enough to guarantee random participant assignment, balanced groups and fully crossed independent variables (Han & Riazi, 2017; Liu et al., 2004). The impact of input rate on SI is influenced by many factors, e.g. the expertise of interpreters (Pio, 2003) and the directionality of interpreting (Chang, 2005), which should be taken into account. Moreover, it is not clear if professional and student interpreters are affected in the same way (Han & Riazi, 2017). Most of these studies chose professional interpreters as participants and only two studies included student interpreters (Liu et al., 2004; Pio, 2003). Besides, no longitudinal studies have been found to explore the effects of input rate on SI fluency at different developmental stages of interpreters' expertise.

Chapter 3 Theoretical Framework

SI involves simultaneous language comprehension and production and is closely related to processes of speech production and the mechanism of memory. This chapter introduces important theoretical models relevant to the current research, including the Effort Model for SI, the model of bilingual speech production and the embedded-processes model of working memory.

3.1 The Effort Model for SI

The Effort Model (Gile, 2009) describes SI as a process consisting of the three core Efforts, namely Listening and Analysis Effort (L), Short-term Memory Effort (M) and Speech Production Effort (P), as well as an additional Coordination Effort (C). The formula for the SI effort model is SI = L + P + M + C. Listening and Analysis Effort is defined as "consisting of all comprehension-oriented operations, from the subconscious analysis of the sound waves carrying the source-language speech which reach the interpreter's ears through the identification of words to the final decisions about the 'meaning' of the utterance" (p. 160). In SI, Production Effort "can be defined as the set of operations extending from the mental representation of the message to be delivered to speech planning and the performance of the speech plan, including self-monitoring and self-correction when necessary" (p. 163). Speech production problems may account for a large part of the fluency phenomena in interpreting. The differences between natural language speaking and interpreting are pointed out by Gile (2009, p. 163): "people are free to speak their own mind and bypass possible production difficulties by rearranging the information and idea sequence, or by dropping or modifying information or using standard phrases. In contrast, interpreters are forced to follow closely the speaker, and waiting for a sentence to finish before being able to start the interpreting would cause excessive short-term memory load". Memory Effort may be required due to the lag between the moment the interpreters hear the speech sounds and the moment they start the interpretation, the time it takes for speech production, individual characteristics of a given speaker or speech such as unclear logic, information density, complex linguistic structure or speaker's accent, or language-specific factors. Coordination Effort corresponds to the resources required to coordinate the three other Efforts.

The Effort Model is based on the single resource theory (Kahneman, 1973), which holds that there is one central pool of resources for different cognitive processing tasks. Demand for the total processing capacity should not exceed the total available capacities of the interpreter and each Effort should have sufficient processing capacity for the task engaged in to achieve smooth interpreting. Processing capacity problems may result in deterioration of the content and/or delivery of the interpreting output (Gile, 2009). Performance deterioration in one or more Efforts may result in errors, omissions, infelicities (EOIs) (Gile, 2011) and disfluencies. Causes of such deterioration of performance (problem triggers) may be cognitive, linguistic or cultural, such as speech density (delivery rate or information density), source speech quality (noise, strong accents, incorrect lexical use etc.), signal vulnerability (numbers and short names), cross-linguistic differences, the speaker's individual speaking style and so on (Gile, 2009).

3.2 The bilingual speech production model

The processes of interpreting production resemble those of speech production and the monitoring system plays similar roles in interpreting and speech production tasks. Interpreting differs from natural language production in that the message is provided by the speaker instead of being conceptualised by the interpreter. The SI process involves multi-tasking in both the comprehension and the production phases and thus has a much higher demand for cognitive coordination than monolingual or bilingual speech production. Models of bilingual speech production provide important references for understanding the interpreting processes.

The well-known framework for L2 speech production is De Bot's (1992) bilingual adaptation of Levelt's (1989) speech production model, which was integrated into the unilingual speaker blueprint by Levelt (1999). The blueprint is important since it provides a summary of what could reasonably be regarded as the consensus view of the linguistic, psycholinguistic and cognitive issues underlying the speaking act (Segalowitz, 2010). Based on Levelt's (1999) blueprint, which integrates De Bot's (1992) observations of L2 speech, Segalowitz (2010) adapted this updated blueprint and specified where L2 fluency issues might arise.

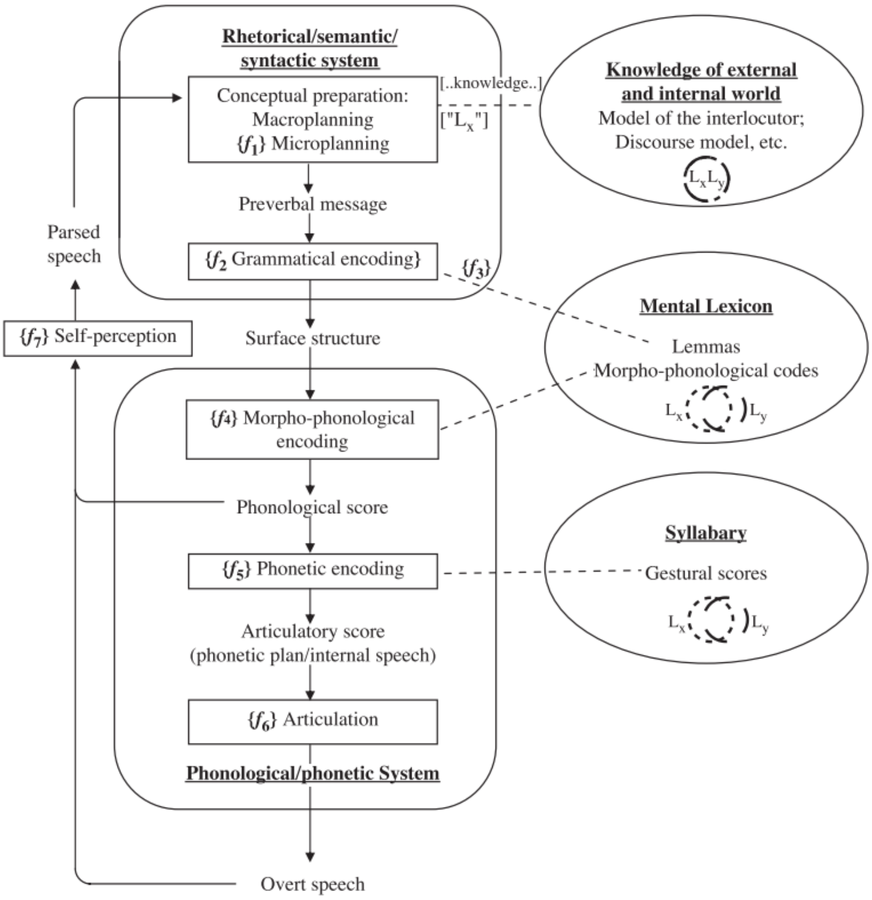


Figure 3.1 The adapted model of L2 speech production (Segalowitz, 2010)

In the above adapted model (Segalowitz, 2010, p. 9), The L_x and L_y circles indicate information pertinent to languages (or registers) x and y. Partially and fully overlapping circles indicate partially distinct and undifferentiated systems, respectively. The {*f*} symbols (fluency vulnerability points) refer to potentially critical points where underlying processing difficulties might be associated with L2 speech disfluencies. All knowledge sources are represented as ellipses.

The first step of speech production is macroplanning (Levelt, 1999), based on the speaker's encyclopaedic knowledge of the external world, knowledge about the interlocutor's internal state of mind (Gallagher & Frith, 2003) and knowledge about discourse conventions. Language choice is determined by the sociopragmatic knowledge of the situation and the language choice information is thus part of the information package along with other encyclopaedic knowledge. It is assumed that the encyclopaedic and social knowledge information is not language-specific (De Bot, 1992; Levelt, 1989, 1999), as represented in the figure by the two completely overlapping circles. There are no L2-specific fluency issues at this macroplanning level since the processes involved are assumed not to be language-specific. But it is pointed out that speech "does not become fluent until the macroplanning process is complete and the systems resources are available solely to speech preparation and production processes" (Roberts & Kirsner, 2000, p. 153).

A subsequent microplanning stage is a process that ends with the formulation of a preverbal message. "Microplanning is far narrower than macroplanning insofar as the content of microplanning output includes only concepts that can be put into words appropriate for the preverbal message" (Segalowitz, 2010, p. 11). For microplanning, the speaker has to decide which information to take into account so that the preverbal message can properly reflect the speaker's construal of the event. A speaker may not know the lexical items needed and figuring out how to build the correct construal information might have a negative impact on fluency (De Bot, 1992). This point of possible vulnerability to disfluency is represented by $\{f_1\}$ at the microplanning level.

The preverbal message is the output of microplanning and may be regarded as a conceptual structure that has not been formulated in words. In order to be represented as the surface structure, the preverbal message must first be encoded into a grammatical structure into which elements from the mental lexicon can be inverted. Grammatical encoding specifies what words to use and how words are related to each other to convey the speaker's intentions and gives linguistic shape to the preverbal message. Difficulty in retrieval and use of appropriate linguistic resources for creating the right grammatical foundation for the surface structure may lead to the possible disfluency vulnerability point identified as $\{f_2\}$ at the grammatical encoding level.

The mental lexicon encompasses all lemmas in each language and must be language-specific to some extent, but there is no need to postulate that there are distinct neural regions in the brain for the speaker's repertoire of different language lemmas (Paradis, 2004). The translation equivalents or near equivalents are assumed to be represented in a neurally similar way. This is why L_x and L_y are partially overlapping in the mental lexicon in the figure. Segalowitz (2010) pointed out that this part of the model is oversimplified, as the overlapping circles miss the distinction between lexicon and vocabulary and separate circles may lead to the misunderstanding that the neural stores for L1 and L2 are separated. A possible disfluency vulnerability point is identified where the grammatical encoding is linked to lemmas in the mental lexicon as $\{f_3\}$, due to the possible difficulty in accessing lemmas during the creation of the surface structure.

Morphophonological codes are associated with lemmas stored in the mental lexicon, making it possible to generate a phonological score required for the generation of overt speech converted from a surface structure. Non-automatic access to syllable programs of the L2 speaker may reduce the fluidity of the process, manifesting itself in hesitations (De Bot, 1992). Thus, another critical point for fluency is identified as $\{f_4\}$ at the morphophonological encoding level, which leads to the formation of a phonological score that underlies the rest of the overt speech formulating processes.

The relatively abstract information in the phonological store is converted into an articulatory score to produce the phonetic acts, which process is called phonetic encoding. Levelt (1999) posits a syllabary, a knowledge source containing the gestural scores for turning phonological score information into motor plans for speech production. The syllabary information includes various parameters for a phonetic event, including local parameters like duration, amplitude, pitch and global parameters like key and register (Levelt, 1999).

The output of the phonetic encoding process is an articulatory score or phonetic plan for starting the motor activity of message articulation. Different repertoires of gestural scores of different languages are used, some highly similar and some quite different across languages as represented by partially overlapping circles in the figure. Fluency issues may appear if the speaker selects the appropriate gestural score and executes that score effortfully instead of automatically, as represented by $\{f_5\}$ and $\{f_6\}$ respectively in the figure.

Speakers in most cases monitor their produced speech at many different levels, as shown in the figure. This self-monitoring may lead speakers to interrupt the fluidity of their speech and reduce speech flow, thus identifying another potential disfluency vulnerability point $\{f_7\}$ as shown in the figure.

In summary, the model points out seven critical points in the speaking system. These fluency vulnerability points represent processing difficulties, which may give rise to L2 disfluency (Segalowitz, 2010). This blueprint enhances our understanding of L1 and L2 speakers and has important implications for fluency issues, but is limited in that it only provides a snapshot of the speaker at one moment in time (De Bot, 1992; Levelt, 1989, 1999). What is missing from the blueprint is an indication of how proficiency skills develop, how speaker-environment interactions influence the act of speaking and the nature of the underlying processes themselves (Segalowitz, 2010).

3.3 The embedded-processes model of working memory

Concurrent speaking and listening in the process of SI imposes great demands on the cognitive resources of interpreters (Chernov, 1979). The construct of working memory serves as "a consistent framework for understanding the cognitive aspects of the complex and demanding skill of simultaneous interpreting" (Liu et al., 2004, p. 20). This section briefly explains the definition and theoretical models of working memory.

Working memory refers to a cognitive system that can temporarily store and process information, which retains information in an accessible state suitable for carrying out tasks with a mental component and is essential for complex cognitive tasks and language processing (Baddeley & Hitch, 1974; Caplan, Waters, & DeDe, 2007; Cowan, 1999). Working memory has a limited capacity and requires "simultaneous storage and processing of information" (Baddeley, 1992, p. 556), and thus plays an essential role in cognitive processing tasks including language comprehension and production. Working memory capacity refers to "attentional processes that maintain task-relevant information activated in an accessible state, or to retrieve that information under conditions of interference, conflict, and competition” (Kane, Conway, Hambrick, & Engle, 2007, p. 23).

There are several different models of working memory (Miyake & Shah, 1999). One of the predominant ones is Baddeley’s (1992, 2000) multicomponent model, which is structural in nature. In this model, working memory is controlled by a limited capacity attentional system, the central executive, aided by two slave systems, i.e. the phonological loop and the visuospatial sketchpad, responsible for the maintenance and temporal storage of acoustic or speech-based information and visual and spatial information, respectively. An additional episodic buffer integrates different types of information and serves as a temporary interface between the slave systems and long-term memory (Baddeley, 2000).

A more suitable model for the current study is Cowan’s (1988, 1999) embedded-processes model of working memory, which is process-oriented and illustrated in Figure 3.2.

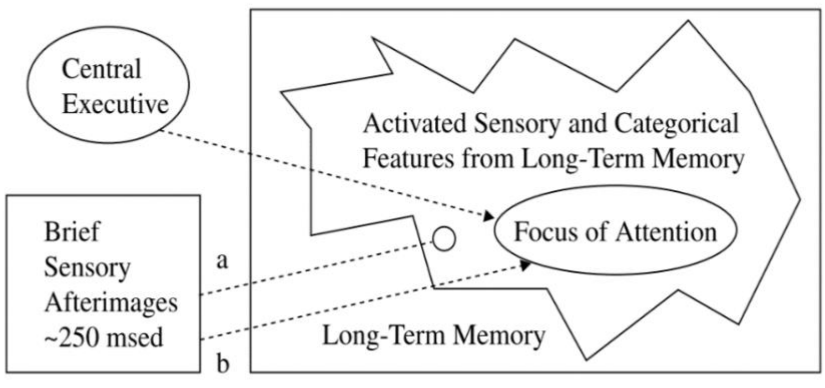


Figure 3.2 Embedded-processes model of working memory (Cowan, 1988, 1999, 2005)

According to Cowan's (1988, 1999, 2005) working memory model, working memory information comes from "hierarchically arranged faculties comprising: (a) long-term memory, (b) the subset of long-term memory that is currently activated, and (c) the subset of activated memory that is in the focus of attention and awareness" (Cowan, 1999, p. 62), as illustrated in Figure 3.2. The large rectangle refers to all information in long-term memory, the jagged shape represents the subset of memory in a temporarily heightened state of activation, and the oval represents the information in the current focus of attention, which is assumed to be a subset of the activated information (ibid., p. 63).

Working memory is a complex construct including "(a) memory in the focus of attention, (b) memory out of the focus but nevertheless temporarily activated, and (c) inactive elements of memory with sufficiently pertinent retrieval cues", and the organisation of working memory is embedded, "with active memory as a subset of long-term memory and the focus of attention as a subset of active memory" (Cowan, 1999, p. 67). Information in the focus of attention is the most readily accessible in working memory. Cowan (1999) argues that the distinction between activation and awareness is important. The information in the focus of attention and possibly all activated information may lead to new links between concurrent or consecutive activated elements and form new composites entered into long-term memory.

The capacity of attention focus is limited to 3-5 unrelated items, though chunking and structure can raise the effective limit; while the activation of memory is time-limited and fades within about 10-20 seconds unless it is reactivated (Cowan, 1997). An important concept related to the current research is retrieval, i.e. entering the correct items into the focus of attention. Items in the focus of attention can probably be recalled. Retrieval from activated memory must occur quickly before the activation fades. When the activated memory representation has disappeared, retrieval of it from the long-term memory is only possible if sufficient episodic memory trace has been stored (Cowan, 1999).

In this model, the regulation of working memory is the control of the focus of attention by the central executive, which is also one way that information in memory can be activated. The allocation of attention is controlled jointly by the involuntary recruitment of attention to especially noticeable events and the voluntary, effort-demanding processes directed by the central executive. The voluntary and involuntary mechanisms work together to determine the focus of attention (Cowan, 1999).

The concept of working memory is central to cognitive psychology as it is assumed to be the vehicle for the retrieval of all information needed to carry out a particular cognitive task (Cowan, 1999). It is also essential for understanding the cognitive bases of interpreting. The embedded-processes model of working memory is a processing framework based on the premise that "the activation mechanisms, attentional and executive mechanisms and long-term retrieval mechanisms all work together in processing to form an effective working memory system" (ibid., p. 97). These mechanisms are important for understanding the interpreting process and the exploration of constructs of cognitive fluency in the current research.

Chapter 4 Research design

This chapter explains the operationalisation of cognitive fluency measures through behavioural experiments in the current research, i.e. the semantic classification task, word translation task, category judgment task and speaking span task. The measurement of utterance fluency indicators and the method of rating perceived fluency are also illustrated. Moreover, the data collection processes are elaborated, including the recruitment of participants, speech manipulation for the SI tasks and the procedures of the behavioural experiments and SI tasks. Finally, the annotation system for the interpreted output is also explained.

4.1 Methodology

Four behavioural experiments were conducted to elicit measures of the cognitive abilities of the participants, i.e. lexical access, lexical retrieval, linguistic attention control and working memory capacity. The behavioural experiments were designed with E-prime 2.0 and presented on a 14-inch ThinkPad laptop.

Measures of utterance fluency in SI performance were obtained through simulated SI tasks simulating conditions of real-life SI. The SI tasks followed a 2 (training: pre/post) * 2 (input rate: low/high) factorial design. Twenty-eight trainee interpreters from MA interpreting programs were recruited as participants. The participants interpreted two speeches, one with a high input rate and the other with a low input speech rate, simultaneously at the beginning and end of an SI training period of one academic term. Three dimensions of utterance fluency were explored, i.e. speed, breakdown and repair fluency. A bilingual corpus of the participants' interpreting output was assembled, with relevant indicators of utterance fluency systematically annotated using Elan 5.2 software. Methods of acoustic analysis were referred to when calculating measures of utterance fluency, for instance, the detection of silent pauses.

Statistical analyses of multiple linear regression and repeated measures ANOVA were conducted with SPSS 24.0 to explore the impact of cognitive fluency measures on the utterance fluency development of the trainee interpreters' SI output, the effects of cognitive fluency, SI training and input rate on utterance fluency in the interpreted production and the relationship between utterance and perceived fluency in the trainee interpreters' SI performance.

4.1.1 Operationalisation of cognitive fluency

Four aspects of the cognitive fluency of trainee interpreters were explored in the current research: the efficiency of lexical access, lexical retrieval, linguistic attention control and working memory capacity. Correspondingly, the four tasks administered were the semantic classification task, the word translation task, the category judgment task and the speaking span task.

4.1.1.1 Lexical access: semantic classification task

Lexical access is the access of lexical entries from the mental lexicon, containing the stored information of the forms and meanings of words, in which basic sound-meaning connections of a language are activated (Field, 2004). It is a fundamental skill required for most aspects of language performance. The semantic classification task was adapted from the design used in the studies conducted by Segalowitz and Freed (2004) and

Segalowitz and Frenkiel-Fishman (2005). In this task, participants made speeded, two-alternative animacy judgments. Single nouns were presented on a computer screen and participants were required to decide whether a word referred to an animate object or not through key responses on Chronos, an external device collecting key or sound responses with millisecond accuracy. The tests were conducted in both L1 (Chinese) and L2 (English) versions.

The English stimulus words were mostly translation equivalents of the Chinese stimuli. Pretests were conducted to ensure that all stimuli in both languages were familiar to bilinguals who had equivalent language competence to that of the participants, i.e. students from the MA translation program. The frequency of stimuli was controlled. The frequency of English stimulus words were checked against the list of the 5,000 most frequently used words according to *A Frequency Dictionary of Contemporary American English: Word sketches, collocates and thematic lists* (Davies & Gardner, 2010). The Chinese stimulus words were chosen from the 5,000 most frequently used characters and words according to the frequency index of *A Frequency Dictionary of Mandarin Chinese: Core vocabulary for learners* (Xiao, Rayson, & McEnery, 2009). The stimulus words used in the task are listed in Appendix 2.

Both L1 and L2 versions of the test began with 30 practice trials, with 15 animate and 15 inanimate stimuli. In the practice trials, participants received feedback on the correctness of their responses after each trial and the results of the practice trials were not included in the final analysis. When the practice task ended, participants pressed "←" to repeat the practice or the "SPACE" key to start the experiment when they were prepared. The experimental procedure included the presentation of 50 animate and 50 inanimate words, recycled twice, leading to a total of 200 experimental trials. In each trial, the participant saw a fixation cross presented on the screen for 150ms. Then a stimulus was presented and would remain on screen for 3000ms until the participant made a key response on the Chronos, followed by a blank screen for 500ms. The order of the stimuli was randomised. The order of task versions (L1 version and L2 version) were counterbalanced across participants. There was a rest after 100 trials and participants pressed the "SPACE" key to continue the experiment. Reaction time and accuracy were recorded.

The efficiency of cognitive processes was operationalised as reaction time speed and stability (coefficient of variability of reaction time) (Ankerstein, 2014; Segalowitz & Segalowitz, 1993). The coefficient of variability (CV) of the reaction time is the standard deviation of an individual's reaction time divided by his or her mean reaction time; "a lower CV reflects more stable reaction times after correcting for the overall speed of responding and, hence, reflects more efficient processing" (Segalowitz & Frenkiel-Fishman, 2005, p. 649), which provides an appropriate index of processing efficiency (automaticity).

4.1.1.2 Lexical retrieval: word translation task

Lexical retrieval involves the selection of lexical concepts which are subsequently encoded, morphologically, phonologically and phonetically, to be either articulated or written down (Levelt, 1989; Snellings et al., 2002). Concepts are transformed into linguistic forms in lexical retrieval; whereas the forms have to be recognised to arrive at the correct lexical concept in lexical access (Snellings et al., 2002). The efficiency of lexical retrieval is an essential subprocess of productive language skills. Previous studies (e.g. Christoffels et al., 2003) have provided evidence for the view that lexical retrieval latencies affect SI performance.

The task chosen to test the efficiency of lexical retrieval in this research was the word translation task, which was adapted from those used in the studies conducted by Christoffels et al. (2003) and De Groot and Poot (1997). Participants were required to say aloud into a microphone the target translation equivalent of the presented word as soon as it appeared on the screen. The translation direction tested was L2-L1 (English to Chinese).

Most stimulus words were chosen from the English word list used by De Groot and Poot (1997). The frequency scores were based on the CELEX frequency count, for which the frequency of English words was based on an English corpus of 18.8 million words (De Groot, Dannenburg, & Van Hell, 1994). The stimulus words were manipulated in terms of frequency, concreteness/imaginability and word length. High frequency and low-frequency stimulus words were included. Words from the two groups were matched based on word length and imaginability. There was no statistically significant difference in terms of word length and imaginability between the high frequency and low frequency group. Different categories of words were included to ensure the generalisability of the findings of the task. The stimulus words of the task are listed in Appendix 3. Pretests were conducted to ensure that all stimuli in both languages were familiar to bilinguals (students of the MA translation program) who had equivalent language competence to that of the participants.

There were 10 practice trials to familiarise the participants with the procedure. In the experimental block, 100 English (L2) stimulus words were presented on the computer screen in a random way. In each trial, a fixation cross was presented on the centre of the screen for 500ms, followed by a 100ms interval of blank screen before the stimulus word appeared in the centre of the screen. The stimulus word stayed on screen for 5000ms until a voice response was detected through a microphone and the Chronos. The reaction time was registered from the onset of the stimulus word. The voice response of the participants was recorded for 2000ms from the moment it was detected. Participants were instructed to remain silent if they did not know the target translation equivalent. Reaction time was registered and response accuracy was checked based on the recordings. CV was calculated as the index of the efficiency of lexical retrieval.

4.1.1.3 Linguistic attention control: category judgment task

Linguistic attention control was operationalised as shift cost, the ability to shift attention between two different attention-directing functions of words, and was measured through the category judgment task. Participants were required to perform both L1 and L2 versions of the task and participants with better attention control were supposed to be able to make such shifts more efficiently.

The category judgment task in the current research was adapted based on previous research (Hu & Wang, 2017; Segalowitz & Frenkiel-Fishman, 2005) and adopted the alternating runs paradigm (García-Amaya, 2012; Hu & Wang, 2017; Rogers & Monsell, 1995). Two sets of stimulus words were used to explore attention-directing functions in this task. One set of stimulus words referred to “the past” and “the future”, which directed the attention of listeners or readers to the temporal location of an event before the present moment (*ago, past, yesterday* and *just now*) or after the present moment (*afterwards, future, tomorrow* and *soon*). The second set of stimulus words involved words of frequency, representing low frequency (*rarely, occasionally, seldom* and *never*) or high frequency (*common, often, frequently* and *always*). Participants were required to judge which category the stimulus words presented on screen belonged to. For the words in the time category, the task was to judge whether the word referred to the past or the future; for the words in the frequency category, the task was to judge whether the word represented low or high frequency. Participants made key responses through Chronos.

Participants received instructions on how to make judgment responses for the time and frequency stimulus words before the task started, followed by four practice blocks of speeded classification trials. The eight time stimulus words in L2 were presented at the centre of the screen randomly and recycled three times, leading to 24 trials in total for Block 1. Block 2 had eight frequency stimulus words in the L2 version. Block 3 and Block 4 were the L1 version of the time and frequency sets of stimulus words, respectively. Each practice block consisted of 24 trials. In each trial of the practice blocks, there was a fixation cross on the screen for 150ms, followed by the stimulus word presented at the centre of the screen. The stimulus would remain on the screen until either the participant made a response or 5000ms had passed with no response. Participants were required to make judgments by pressing the response keys on the Chronos as quickly as possible. There was feedback on whether the response was correct or not after each trial. After each block, feedback on the error rate and mean reaction time for that block was presented on the screen. Participants could choose to repeat the practice for that block or press the spacebar to continue with the next block.

The tasks were administered in both L1 (Chinese) and L2 (English). The stimulus words in the Chinese version were mostly the translation equivalents of those in the English version. The stimulus words used in the task are listed in Appendix 4. Eight L1 blocks and eight L2 blocks alternated, constituting 16 blocks in total. The order of the language of blocks was counterbalanced: half of the participants completed the 16 blocks in the "L1L2L1L2…" order and the rest in the "L2L1L2L1…" order. The L1 and L2 blocks were distributed evenly across the session.

Within each block, the two judgment tasks, i.e. time and frequency, alternated. The time (T) and frequency (F) words were presented in the sequence "…TTFFTTFFTTFF…", thus alternating between repeating and shifting conditions in a predictable way. Stimulus words from the time or frequency word sets were presented randomly, two adjacent words not being repeated. In each block, the eight "time" words and the eight "frequency" words were repeated three times, leading to a list of 48 stimulus words. Studies adopting the alternating runs procedure have found that people perform faster on repeat trials than on shift trials, even if the participants can predict which trial is upcoming (Monsell, Sumner, & Waters, 2003; Rogers & Monsell, 1995; Wylie & Allport, 2000). The difference between repeating and shifting conditions is referred to as shift cost, which reflects the extra burden the processing system carries in order to change the focus of attention (Segalowitz & Frenkiel-Fishman, 2005).

In each trial, stimuli were presented clockwise in the four quadrants of a square (10cm*10cm) in the middle of the screen. Each stimulus word was presented at the centre of one quadrant each time. The quadrants in which the first stimulus appeared was randomised across participants, which meant that the first stimulus word might appear in any of the four quadrants. In the subsequent trial, a new stimulus appeared in the adjacent quadrant to that of the previous trial, moving clockwise around the screen. Positions of a stimulus word served as visual cues as to which task (time or frequency judgment) was to be performed. In the experimental stage, each stimulus word would stay on the screen until either the participant made a response through the Chronos keys or 5000ms had passed with no response. The response-stimulus interval was 150ms. There was visual feedback for 20ms when the response was incorrect. In the case of an incorrect response, the stimulus-response interval was prolonged for an additional 1,500ms to allow participants to recover, and data from the incorrect trial and the subsequent one were discarded (Rogers & Monsell, 1995). The error rate and mean reaction time were presented on the screen after each block. There was a rest every four blocks.

The mean reaction time and CV of the repeating and shifting conditions in each language version were calculated. The shifting cost indexes in the current research were calculated as the measures of CV under shifting conditions minus the corresponding measures under repeating conditions.

4.1.1.4 Working memory capacity: speaking span task

A speaking span task (Daneman, 1991; Daneman & Green, 1986) was chosen to test the working memory capacity of participants in this research. It is worth pointing out that reading span tasks and their variants (listening and speaking span tasks) are associated with language processing in particular (Wen, 2012). The speaking span test taxes the processing and storage of memory simultaneously during the production process (Daneman & Green, 1986). It is a variant of the reading span task (Daneman & Carpenter, 1980). However, speaking span is found to be related to verbal fluency in both speech and reading tasks (Daneman, 1991). The tests were administered in both L1 and L2 versions since memory capacity may be different in native and second languages (Chincotta & Underwood, 1998; Service, Simola, Metsänheimo, & Maury, 2002). The stimulus words used in the task are listed in Appendix 5.

The speaking span task in the current research followed previous research (Christoffels et al., 2003; Daneman & Green, 1986; Fortkamp & Bergsleithner, 2007; Guará-Tavares, 2013; Mota, 2003; Jin, 2012). Sixty unrelated English (L2) words were selected. All stimulus words were high-frequency seven-letter words chosen from *Collins COBUILD Learner's Dictionary* (marked five points in terms of word frequency). The stimulus words were individually presented in the middle of the computer screen using E-prime software for 1000ms, followed by a 500ms blank screen before the next stimulus word appeared. The words were presented in three series and each series contained 20 words. In each series, the two-word set was presented first, followed by the three-, four-, five- and six-word sets consecutively. Words within each set were not related, both semantically and phonologically, to avoid the use of strategies by participants when memorising the presented words. Participants were asked to read each word silently and remember the words. At the end of each set, a visual signal (question marks) appeared on the screen with an accompanying tone to signal the end of the set. The number of question marks equalled the number of words in the set just presented. Participants were required to generate verbally a set of grammatically acceptable sentences (both semantically and syntactically) for each of the words just presented in the original order and form. There were no restrictions of the length and complexity of the produced sentences, or the position of the recalled words in the sentences. When the participant had finished the recall and production of sentences for the current set, the next set of words were triggered until all 60 words had been presented. Participants familiarised themselves with the task in practice trials before starting the experimental trials.

The scoring for speaking span can be strict or lenient. The two calculation methods differ in that the former requires the exact original forms of the presented words used when producing the required sentences and the latter accepts derivative forms. The scoring for the speaking span task in the current research adopted the strict calculation method, for which the Strict Speaking Span (SSS) and the Composite Strict Speaking Span (CSSS) are the candidate indexes (Jin, 2011; Li & Yu, 2009).

The Strict Speaking Span (SSS) is the total number of words for which grammatically and semantically acceptable sentences are produced using the exact form of the presented words, but it does not require the order to be the same as the presented order (Daneman, 1991; Daneman & Green, 1986). The maximum score is 60. It has been argued that the total performance score could approximate continuous variables like working memory span better (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). The Composite Strict Speaking Span (CSSS) takes three dimensions into consideration, i.e. processing accuracy, processing efficiency and storage ability, and has been proven to be able to better reflect the functions of working memory and predict oral fluency (Jin, 2011, 2012; Weissheimer, 2007). It has been argued that the traditional measurement of speaking span, which only measures the storage capacity, might not reflect differences in processing efficiency and storage ability (Jin, 2011). The calculation of CSSS in the current study follows that in Jin’s (2011, 2012) studies. Processing accuracy is calculated as the number of syntactically and semantically acceptable sentences produced with the original form of presented words, not requiring the serial order of words. Processing proficiency is the ratio of the time used to produce these sentences to the total number of sentences, reflecting the average reaction time. When scoring processing proficiency, words recalled in the original order score 1 point each, otherwise 0.5 points each. The average reaction time of correct responses is multiplied by -1. Storage ability is operationalised as the overall scores of words recalled, including those in incorrect sentences, in derivative forms, or words recalled without formulating sentences. The Composite Strict Speaking Span (CSSS) is the average of the above three items after standardisation and was adopted as the index for speaking span in the current research.

4.1.2 Measures of utterance fluency

Researchers have been inconsistent in the way they operationalise oral performance variables of fluency. For example, they have used different thresholds and methods of measuring silent pauses, made different choices regarding syllables or words per second or minute when measuring speech rate, used different sample sizes and speech elicitation techniques and so on (Kormos, 2006). Another reason for the lack of convergence on oral indicators of fluency is that oral performance measures may reflect more than one function (Segalowitz, 2000).

Previous research in oral fluency has mostly used temporal variables related to fluidity of output and performance variables related to language expression and disfluencies (Foster, Tonkyn, & Wigglesworth, 2000; Freed, 1995; Kormos, 2006; Kormos & Dénes, 2004; Lennon, 1990; Riggenbach, 1991; Towell et al., 1996; Zhang & Wu, 2001). Five temporal measures were used in the study conducted by Towell et al. (1996): speech rate, average length of pauses, phonation time ratio, mean length of run and articulation rate. Lennon's (1990) study investigating fluency in EFL included 12 temporal measures and disfluency markers (repetitions/T-Unit, self-corrections/T-Unit, filled pauses/T-Unit and percentage of repeated and self-corrected words). Kormos (2006) summarised that most previous studies concluded that speech rate, mean length of run and phonation time ratio were the best predictors of fluency. Zhang and Wu (2001) included in their study indexes for content coherence (ratio of reported necessary events to total necessary events) and indexes for language acceptability (ratio of error-free T-units to total T-units to measure accuracy, mean length of c-units after pruning and subordinate clauses per T-unit to measure syntactic complexity). In total, four categories of 12 fluency measures were used, including temporal, content, linguistic and performing indexes. Mead (2005) introduced five temporal variables: speech rate, duration of pauses, phonation time ratio (percentage of speaking time, as opposed to pauses), articulation rate (total word count divided by phonation time) and mean length of run (mean length of speech segments uninterrupted by pauses). He pointed out that three of the parameters (speech rate, pause duration and mean length of run) might be enough for practical purposes as phonation time ratio and articulation rate offered different perspectives on the same data.

In the current research, measurements of the three dimensions of utterance fluency (Skehan, 2003; Tavakoli & Skehan, 2005), i.e. speed fluency, breakdown fluency and repair fluency, are included, following the measurement of utterance fluency in previous research on second language production and interpreting (Bosker et al., 2012; Foster & Skehan, 1996; Han, 2015b; Kormos, 2006; Kormos & Dénes, 2004; Lennon, 1990; Mead, 2005; Yu & van Heuven, 2017).

The threshold duration for unfilled pauses is important as it affects the calculation of most of the fluency indicators. The minimum duration chosen for unfilled pauses in previous research is inconsistent. The threshold commonly used is 0.25 seconds (Cecot, 2001; Duez, 1982; Goldman-Eisler, 1958; Tissi, 2000) or 0.3 seconds (Raupach, 1980; Wang & Li, 2015). Some studies chose to measure longer silent pauses that were perceived as disfluency, e.g. 0.4s in Freed's (2000) study. The threshold of 0.3 seconds is regarded as the minimum duration for a silent pause in the current research.

The tempo of interpreters' output in SI is to a large extent paced by the speaker. The calculation of measures of utterance fluency in SI should take into account long pauses in original speeches. Following the example of Pöchhacker (1995), measures of utterance fluency indicators in the current research were adjusted for extended pauses in the source speeches in order to obtain a realistic indication of fluency measures. All silent pauses longer than 2 seconds in the source speeches were annotated and the total duration of these long silent pauses was subtracted from the total time duration of the interpreting output when calculating the utterance fluency measures. Indexes of utterance fluency measured in this research are summarised in Table 4.1.

Table 4.1 Indicators of SI utterance fluency in the current research

Dimensions	Indicators	
Speed fluency	speech rate (SR) articulation rate (AR) mean length of run (MLR) phonation time ratio (PTR)	
Breakdown fluency	silent pauses (SP)	mean number of silent pauses mean duration of silent pauses
	filled pauses (FP)	mean number of filled pauses mean duration of filled pauses
Repair fluency	REPAIRs	mean number of repairs, repetitions and false starts mean duration of repairs, repetitions and false starts

Notes: mean number is calculated as the mean number per minute

4.1.2.1 Measures of speed fluency

Four temporal variables were measured for speed fluency in SI performance in this research: speech rate (SR), articulation rate (AR), mean length of run (MLR) and phonation time ratio (PTR). Since the target language was Chinese in this research, the calculation was based on the total number of Chinese characters for the interpreting output. All measures were adjusted for extended pauses (≥2s) in the source speeches. A major methodological issue related to the measurement of MLR is deciding if it is delimited only by silent pauses or by silent and filled pauses together (Mead, 2005; Towell et al., 1996). In this research, only silent pauses were used to delimit a run. Measures of speed fluency were defined and calculated as follows, based on previous research (Kormos, 2006; Lennon, 1990; Mead, 2005; Towell et al., 1996; Yu & van Heuven, 2017).

Methods of calculation for speed fluency measures in this research:

1) SR: speech rate refers to the total number of words or characters produced, including disfluencies, divided by the total duration of speech (including pauses);
2) AR: articulation rate is the number of words or characters, including disfluencies, divided by the total duration of speech apart from all (silent and filled) pauses longer than 0.3 seconds;
3) MLR: mean length of run is the number of words or characters in utterances between pauses of 0.3s and above;
4) PTR: phonation time ratio is the percentage of speaking time divided by the total speech time.

4.1.2.2 Measures of breakdown fluency

Breakdown fluency is related to pauses, which have been proven to play an important role in fluency (Cucchiarini et al., 2002; Freed, 1995; Lennon, 1990; Pradas Macías, 2006; Raupach, 1987; Riggenbach, 1991). Filled and unfilled (silent) pauses are distinguished from each other and filled pauses include fillers and lengthenings of syllables or words in this research.

Measures of breakdown fluency investigated in the current research were the mean number and duration of silent pauses and mean number and duration of filled pauses. All measures were adjusted for extended pauses (≥2s) in the source speeches.

Methods of calculation for breakdown fluency measures in this research:

1) SP mean: the mean number of silent pauses is the total number of silent pauses per minute, calculated as the total number of silent pauses divided by the total amount of speaking time, expressed in seconds and multiplied by 60;

2) SP length: the mean length of silent pauses refers to the mean duration of silent pauses longer than 0.3 seconds;
3) FP mean: the mean number of filled pauses is the total number of filled pauses per minute, calculated as the total number of filled pauses divided by the total amount of speaking time (adjusted for extended pauses in the source speeches), expressed in seconds and multiplied by 60;
4) FP length: the mean length of filled pauses refers to the mean duration of filled pauses, as annotated using Elan software.

4.1.2.3 Measures of repair fluency

Measures of repair fluency investigated in this research include the mean number and mean length of REPAIRs, i.e. the sum of repairs, repetitions and false starts. Repairs, repetitions and false starts in the interpreted output were annotated using Elan software, which provided the statistical information regarding their amount and duration. All measures were adjusted for extended pauses (≥2s) in the source speeches.

Methods of calculation for repair fluency measures in this research:

1）REPAIR mean: the mean number of REPAIRs is the total number of repairs, repetitions and false starts per minute, calculated as total number of repairs, repetitions and false starts divided by the total amount of speaking time (adjusted for extended pauses in the source speeches), expressed in seconds and multiplied by 60;
2）REPAIR length: the mean length of REPAIRs refers to the mean duration of repairs, repetitions and false starts.

4.1.3 Operationalisation of perceived fluency

4.1.3.1 Raters

The perception of fluency in the SI output of trainee interpreters was operationalised as ratings of a homogenous group of human raters (Cheung, 2007, 2013; Hamidi & Pöchhacker, 2007; Lee, 2008; Liu, 2013). The 112 recordings of SI performance (4 recordings for each of the 28 participants) were rated by three raters. The use of multiple raters is believed to reduce uncertainties of measurement and inconsistencies across raters and contribute to rating reliability (Han, 2018). All three raters, with the average age of 33 years old, were native Chinese speakers with excellent command of English and had professional interpreting experience. Two raters (both female, aged 37 and 35 respectively) were full-time interpreting teachers with professional interpreting experience. They had experience teaching interpreting at the undergraduate and postgraduate levels in universities for about five years and experience assessing interpreting performance for CATTI (China Accreditation Test for Translators and Interpreters). The third rater (male, aged 28) was a professional interpreter with two years of professional SI experience and part-time interpreting teaching experience of one year.

4.1.3.2 Rating scale

When assessing interpreting performance, three main criteria of SI performance, i.e. content, delivery and language quality, were frequently documented in previous studies (e.g. Han & Slatyer, 2016; Lee, 2008; Liu, 2013). Raters were asked to rate fluency, which was the focus of the current research, and overall interpreting performance of trainee interpreters, a relevant aspect in the current research. Rating of overall SI quality was included to make further investigation possible. More detailed aspects of fluency such as control of pauses or disfluencies were not rated to avoid the Halo effect. Raters might find it difficult to differentiate different aspects of rating criteria, which could result in highly correlated scores (Han, 2018).

Descriptor-based rating scales were constructed to assist the rating process, and contained five bands: 90-100, 80-89, 70-79, 60-69 and 0-59. No participants received a rating within the rank 0-59 for either fluency or overall interpreting performance. Details of the rating scales used in the rating process are listed in Appendix 16. The two scales were for fluency and overall interpreting performance respectively. Descriptors of the scales were adapted from scales in previous studies (e.g. Han, 2016; Lee, 2008).

4.1.3.3 Rating procedure

For the simulated interpreting tasks, 28 student interpreters were asked to interpret four speeches: two speeches in the pre-training task and two speeches in the post-training task, one with a low input rate and the other with a high input rate for both tasks. The interpreting performance was recorded, which led to 112 (28*4) audio files in total. The interpreting files were distributed to each rater in a random order to avoid the order effect, a potential threat to the consistency with which rating criteria are applied (Bachman & Palmer, 1996).

Before the rating process started, raters received the same amount of individual training from the researcher. The researcher explained the goal and context of rating, the construct and content of the simulated interpreting tasks, and the rating criteria. The concept of utterance fluency to be rated was explained and made clear to all raters. In the training session, the raters listened to several sample interpreting recordings as a practice and discussed the rating results with the researcher.

The rating session was post-hoc and the raters completed the rating session individually at their convenience. Post-hoc rating (Han, 2015b; Lee, 2008; Liu, 2013; Meuleman & Van Besien, 2009; Wang & Napier, 2015) was adopted since geographical constraints and the amount of work made group rating on-the-spot impossible. The source speech texts and reference translations were provided for raters to get familiar with the content of source speeches and help to check the accuracy of interpreting output when rating overall interpreting performance. The official Chinese versions of the speeches provided by the Prime Minister's Office of Singapore were used as reference translations. Raters evaluated each performance on both scales (fluency and overall interpreting performance) in one listening instead of listening to the interpretation twice. After the raters completed their individual ratings, a weighted average was calculated to obtain an overall score for each recording (Han, 2015b; Lee, 2008).

4.2 Data collection

4.2.1 Participants

Twenty-eight trainee interpreters from MA interpreting programs from three universities in Hong Kong, 26 female and 2 male, were recruited as participants using convenient sampling method. Background information of participants was surveyed through questionnaire (Appendix 1). They all had Chinese as A language and English as B language, except one participant who was a natural bilingual. Their mean age was 23.7 years old (SD = 1.2). IELTS score is used as the index for general English proficiency (mean score = 7.4, SD = 0.4). The participants had on average received 1.5 years of consecutive interpreting training (about 1 year at the undergraduate level and 0.5 years at the MA level), and by the time the experiment started, they were all trainee interpreters at the beginning stage of SI training. Participants received compensation for completing all experiment sessions. The training the participants received was comparable across the three universities. The participants received a three-hour SI classroom working session each week. The interpreting teachers guided participants in their development of SI strategies and skills during the classroom session, designed and provided participants with interpreting materials for their personal practice after class. The average SI practice time for participants was about 15-20 hours each week during the period the research was conducted.

4.2.2 Speech manipulation for the SI tasks

The four source speeches interpreted, two for the pre-training tasks and two for the post-training tasks, were adapted from authentic speech videos. To ensure the comparability of the four speeches, all speeches were delivered by the same speaker, the Prime Minister of Singapore, Lee Hsien Loong, for the National Day Rally, an annual event. Four speeches were adapted from four National Day Rally speeches on general topics and the speed of the speeches was manipulated to produce one slower speed (S) version (about 120 words per minute) and one faster speed (F) version (about 140 words per minute) for each speech. The adaptation and manipulation of speech videos were realised using Corel VideoStudio Pro X10 software, which allows speed adjustment and pitch shift. The pitch of eight manipulated speeches was maintained at the same level after adjustment. Information regarding the eight manipulated speeches is listed in Table 4.2.

Table 4.2 Source speech information on the SI tasks

Speeches	Pre-training				Post-training			
	Speech A		Speech B		Speech C		Speech D	
	S	F	S	F	S	F	S	F
Word count	1534		1558		1555		1563	
Duration(min)	13.02	10.92	13.18	11.08	13.25	11.03	13.18	11.08
Speed (wpm)	117.82	140.48	118.21	140.61	117.36	140.98	118.59	141.06

Notes: S for speeches of low input rate; F for speeches of high input rate

Efforts were made to ensure that the adapted speeches were linguistically comparable. A set of lexical, syntactic and discourse parameters were derived to ascertain the comparability using Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; Hild, 2011), a computer tool for analysing texts on diverse measures of cohesion, language and readability. Latent semantic analysis (LSA), a statistical representation of word and text meaning is adopted in Coh-Metrix as a measure of semantic cohesion and coherence (Foltz, 1996; Landauer & Dumais, 1997). Represented indexes derived are listed in Table 4.3. Though it was impossible to ensure all the linguistics indexes to be the same, efforts were made to ensure key linguistic features of the four speeches were comparable and did not differ to a large extent. Moreover, two professional interpreters were invited to listen to the adapted speeches and confirmed that they were natural for interpreters.

Table 4.3 Text indexes of source speeches

Analysis level	Indices	Pre-training		Post-training	
		Speech A	**Speech B**	**Speech C**	**Speech D**
Lexical	type-token ratio (lexical density)	0.58	0.58	0.55	0.6
	word length (mean syllable number)	1.59	1.53	1.51	1.54
Syntactic	NP density (mean modifier number per noun phrase)	0.67	0.7	0.74	0.74
	Flesch reading ease	58.31	62.77	62.76	63.14
LSA	given/new, sentences, mean	0.28	0.3	0.3	0.29
	overlap, adjacent sentences, mean	0.11	0.13	0.14	0.13

4.2.3 Experimental procedures

The experiment was conducted alongside an intensive, thirteen-week SI training period in which the participants engaged. Participants were tested individually in three sessions of the experiment: in session one, the explored cognitive abilities of participants were tested through four behavioural tasks (three in both L1 and L2 versions) at the beginning of SI training; in session two, the first simulated SI task was conducted at the starting stage of SI training; and in session three, the second simulated SI task was conducted at the end of SI training. An overview of the three sessions is listed in Table 4.3.

4.2.3.1 Procedure of behavioural experiments

Before the experiment started, the participants were presented with an information sheet regarding the research, introducing the purpose of the research, ethical approval, main procedures, potential benefits and risks, the issue of confidentiality and so on.

Participants were encouraged to ask questions if they had any doubts. Participants then signed the consent form to participate in the experiment voluntarily and they understood that they could withdraw at any time and at any stage of the experiment. Participants' personal data were collected, including demographic information, educational background, interpreting experience and IELTS score.

For the behavioural experiments, participants sat in a quiet environment and responded to the stimuli presented on a laptop based on the instructions. Four cognitive tasks, i.e. the semantic classification task to test the speed and efficiency of lexical access (LA), the word translation task to test the speed and efficiency of lexical retrieval (LR), the category judgment task to test the flexibility of linguistic attention control (AC) and the speaking span task (SS) to test working memory capacity, were conducted in session 1 at the beginning of SI training. Three of the four tasks, i.e. the semantic classification task, category judgment task and speaking span task, were administered in both L1 (Chinese) and L2 (English) versions. Practice trials preceded the experimental trials to prepare the participants. The order of different tasks and of L1 and L2 versions for each task were counterbalanced among the participants. The tasks were presented on a 14-inch ThinkPad laptop screen using E-prime 2.0 software (Schneider & Zuccoloto, 2007). Participants responded through Chronos and a microphone.

4.2.3.2 Procedure of SI tasks

The simulated SI tasks were conducted at two time slots, i.e. the start and the end of the SI training period. The participants individually interpreted two speeches simultaneously in a sound-proof SI booth in each session. The SI tasks simulated a real conference environment as much as possible. A small audience listened to the interpreting on site. The speech videos were played on the computer screen in the booths. The speeches were adapted from live speeches, which were made by the same speaker and on the same occasions, with different but comparable content. The four speeches were manipulated with Corel VideoStudio X10 software into one slower version of about 120 words per minute and one faster version of about 140 words per minute, as explained in section 4.2.2. The participants' interpreting performance was recorded digitally with a double-track recording system.

In session 2, the first round of SI tasks, the participants interpreted speech A and speech B. Each participant interpreted a slower version of one speech and a faster version of the other. The order of speeches (A or B) and the order of input rate versions (slow or fast) were counter-balanced using Latin-square design. Twenty-eight participants were randomly grouped into the following four groups in a counter-balanced way: Speech A_Slow, Speech B_Fast; Speech B_Fast, Speech A_Slow; Speech A_Fast, Speech B_Slow; Speech B_Slow, Speech A_Fast, with seven participants in each group.

For each speech, participants familiarised themselves with the SI equipment before the experiment started. Each participant controlled the equipment themselves during the interpreting tasks. A briefing note of the topic of the speech, background information on the speech and the speaker and a glossary were distributed to the participants about ten minutes in advance to aid their preparation. A warm-up speech made by the same speaker and on the same occasion allowed the participants to get familiar with the speaking style of the speaker. The warm-up video lasted for about 4 minutes and was delivered at a delivery rate of about 130 words per minute. The participants only started the subsequent interpreting task once they were ready. After interpreting the first speech, participants filled out a questionnaire to rate the level of difficulty of the source speech and their SI performance in terms of content and fluency. The questionnaire for difficulty assessment and self-evaluation of performance is in Appendix 7. Participants were required to have a break of at least ten minutes after finishing the first questionnaire to avoid the effects of fatigue. The procedure for interpreting the second speech was the same as that for the first speech.

The second round of SI tasks (session 3), which was conducted at the end of the SI training period, ran the same procedure as that of the first round. In this session, participants interpreted speech C and speech D, which were made by the same speaker and comparable to speech A and speech B. The order of speeches (C or D) and the order of input rate versions (slow or fast) were counter-balanced in the same way.

Table 4.4 Experimental procedures

Session 1

Introduction	briefing of the research goal, procedures etc.	
Consent form		
Participants information	demographic data, educational background, IELTS score etc.	
Cognitive tasks	LA	lexical access/semantic classification task (C&E)
	LR	lexical retrieval/word translation task (E-C)
	AC	linguistic attention control/category judgment (C&E)
	SS	speaking span task (C&E)

Session 2

Introduction	Description of the procedures of the SI task
Equipment training	in SI booth
Briefing note	background information for the event and speaker, glossary
Warm-up practice	participants start once they feel ready
SI for speech A	order of speeh A&B and version S&F counter-balanced
Questionnaire A	Participants rate source speech difficulty and their SI performance

Break	At least 10 minutes
SI for speech B	the same procedure as that for speech A
Questionnaire B	

Session 3

Introduction	Order of speech C&D and version S&F are counter-balanced; other procedures are the same as in Session 2
Equipment training	
Briefing note	
Warm-up practice	
SI for speech C	
Questionnaire C	
Break	
SI for speech D	
Questionnaire D	

4.3 Annotation

The source speeches and respective interpreting renditions were transcribed and synchronised with oscillograms using Elan 5.2 software, a professional tool for the creation of complex annotations on video and audio resources. The application of Elan software in interpreting research has been done in previous research, e.g. Mizuno (2017). Elan can convert acoustic signals into an oscillogram and visualise the sounds like a continuous wave pattern, offering statistics regarding the frequency and duration of annotations. Through synchronisation with the oscillogram timeline, the annotation markers can be accurate to one millisecond.

Table 4.5 The annotation system for the SI output

Tier	Annotation Type	Annotated Content
1	Text	texts
2	Pauses	silent pauses (SP)
		filled pauses (FP)
3	REPAIRs	repairs
		repetitions
		false starts

Table 4.5 illustrates the annotation system for the SI output. Multiple tiers of annotations were created: Tier 1 for texts of the SI output; Tier 2 for silent and filled pauses, which were indicators of breakdown fluency measures; and Tier 3 for indicators of repair fluency measures (REPAIRs), including repairs, repetitions and false starts. For the source speeches, only the texts were annotated since there were rare disfluencies. Tiers of the source speeches and their SI output could be merged within the software to create a synchronised version when necessary.

The threshold of 0.3s for an unfilled pause was adopted. An annotation refers to one run of words between two unfilled pauses (≥0.3s) when annotating texts. The annotations were coded twice (one month apart) by the researcher and high intra-rater reliability was guaranteed (Liu, 2011). The following is an example of annotation in the Elan software.

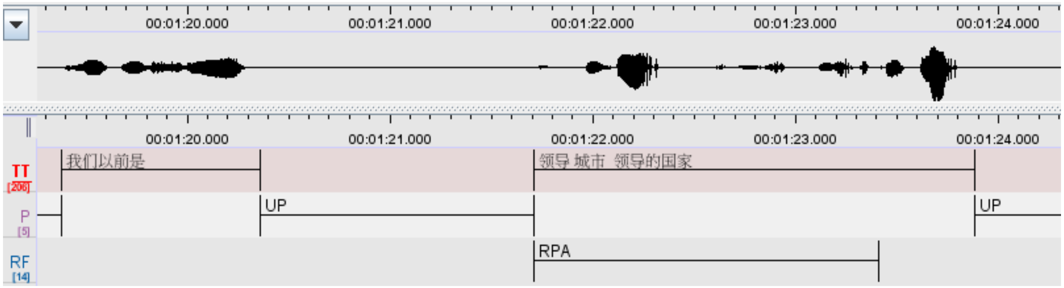


Figure 4.1 An annotation example in Elan

Notes: Tier 1 (TT) is the transcription of the SI output, synchronised with the timeline; Tier 2 (P) marks silent and filled pauses; Tier 3 (RF) annotates repairs, repetitions and false starts in the SI output.

The SI tasks produced a total of 224,369 Chinese characters in the interpreted output, leading to 29,253 annotations of the SI texts. Moreover, 29,141 silent pauses (SP), 1,020 filled pauses (FP), 1,364 repairs, 532 repetitions and 324 false starts (FS) were annotated.

Table 4.6 The number of annotations in the SI output

	SI TEXTS		**PAUSES**		**REPAIRs**		
	Count	**Annotations**	**SP**	**FP**	**Repairs**	**Repetitions**	**FS**
Pre_S	58,302	7,947	7,919	274	339	143	95
Pre_F	52,897	6,731	6,703	310	372	160	63
Post_S	59,531	7,917	7,889	227	372	122	81
Post_F	53,639	6,658	6,630	219	281	107	85
Sum	224,369	29,253	29,141	1,030	1,364	532	324

Chapter 5 The role of cognitive fluency in SI utterance fluency development

A series of multiple linear regression analyses were conducted, using SPSS 24.0. Standard multiple regression, also referred to as simultaneous regression, was selected as the method of regression analyses to explore the influence of cognitive fluency measures on the development of utterance fluency in trainee interpreters' SI output. Standard multiple regression "is primarily useful for explanatory research to determine the extent of the influence of one or more variables on some outcome" (Keith, 2015, p. 80). It "estimates the direct effects of each independent variable on the dependent variable" and can "be used to determine the extent to which a set of variables predicts an outcome and the relative importance of the various predictors" (ibid.). In simultaneous regression, all independent variables are entered into the regression simultaneously; the respective contribution of each independent variable is assessed as if it had been entered into the regression after all other independent variables had been entered; and the R_2 of simultaneous regression is the sum of the unique contributions made by each of the predictor variables (Plonsky, 2015; Tabachnick & Fidell, 2012).

The standard multiple linear regression analyses were performed taking measures of cognitive fluency as the independent variables and measures of utterance fluency development (partialling out measures of utterance fluency in the pre-test from those in the post-test) under conditions of low and high input rate as individual dependent variables, listed in Table 5.1 below.

Table 5.1 List of independent and dependent variables of standard multiple linear regression analyses

Independent Variables (Cognitive Fluency)	Lexical access	LA EN	The CV measures for the English version of the lexical access task
		LA CH	The CV measures for the Chinese version of the lexical access task
	Lexical retrieval	LR	The CV measures for the lexical retrieval task (E-C)
	Linguistic attention control	AC EN	The CV measures for the English version of the linguistic attention control task
		AC CH	The CV measures for the Chinese version of the linguistic attention control task
	Working memory capacity	SS EN	CSSS for the English version of the speaking span task
		SS CH	CSSS for the Chinese version of the speaking span task
Dependent Variables (Development of Utterance Fluency)	Speed fluency	SR	change in speech rate
		AR	change in articulation rate
		PTR	change in phonation time ratio
		MLR	change in mean length of run
	Breakdown fluency	SP mean	change in mean number of silent pauses per minute
		SP length	change in mean duration of silent pauses
		FP mean	change in mean number of filled pauses per minute
		FP length	change in mean duration of filled pauses
	Repair fluency	REPAIR mean	change in mean number of REPAIRs (repairs, repetitions and false starts) per minute
		REPAIR length	change in mean duration of REPAIRs (repairs, repetitions and false starts)

Notes: CV for coefficient of variance of reaction time; CSSS for composite strict speaking span

Results of standard multiple linear regression analyses are presented and discussed in the following. The impact of cognitive fluency on the development of the three dimensions of utterance fluency, i.e. the speed, breakdown and repair fluency measures, under conditions of low and high input rate in trainee interpreters' SI output are discussed, respectively.

5.1 Data analysis

5.1.1 Influence of cognitive fluency on speed fluency development

For the development of speed fluency in the SI output, results of the multiple linear regression analyses showed that the cognitive fluency measures had a significant impact on SR (speech rate), AR (articulation rate), PTR (phonation time ratio) and MLR (mean length of run) under conditions of high input rate. Besides, the cognitive fluency measures did not affect SR, AR, PTR, MLR under low input rate conditions significantly. Results of analyses with significant results are presented in the following.

5.1.1.1 Influence on speech rate

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in SR as an indicator of the development of speed fluency under conditions of low and high input rate, as SR was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on S_SR (change in speech rate under low input rate conditions) was not significant. Results of the analysis with F_SR (change in speech rate under high input rate conditions) as the dependent variable are shown in Table 5.2.

Table 5.2 Coefficients of the regression model for F_SR

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
(Constant)	.187	.218		.857	.402		
LA_EN	2.317	.902	.764	2.570	.018	.294	3.405
LA_CH	-1.894	.874	-.685	-2.166	.043	.260	3.852
LR	-.664	.526	-.224	-1.261	.222	.820	1.220
AC_EN	-.427	.371	-.197	-1.150	.264	.882	1.134
AC_CH	.915	.447	.370	2.047	.054	.794	1.260
SS_EN	.130	.072	.415	1.813	.085	.496	2.016
SS_CH	.117	.069	.349	1.706	.104	.620	1.613

Results of the multiple linear regression analysis showed that the linear regression model was moderately satisfactory, the goodness of fit reaching 48.1%. Results of T-tests for the regression coefficients showed that LA EN (t=2.570, p=0.018<0.05) and LA CH (t=-2.166, p=0.043<0.05) affected F_SR significantly. The impact of LA EN on F_SR was significantly positive with the coefficient 2.317 > 0, indicating that F_SR was larger when LA EN was larger. The impact of LA CH on F_SR was significantly negative with the coefficient -1.894 < 0, indicating that F_SR was larger when LA CH was smaller. Besides, other variables like LR, AC EN, AC CH, SS EN and SS CH did not affect the dependent variable F_SR significantly.

5.1.1.2 Influence on articulation rate

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in AR (articulation rate) as an indicator of the development of speed fluency under conditions of low and high input rate, as AR was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on S_AR (change in articulation rate under low input rate conditions) was not significant. Results of the analysis with F_AR (change in articulation rate under high input rate conditions) as the dependent variable are as follows.

Table 5.3 Coefficients of the regression model for F_AR

Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	-.251	.272		-.923	.367		
LA_EN	1.367	1.127	.369	1.213	.239	.294	3.405
LA_CH	-.189	1.093	-.056	-.173	.864	.260	3.852
LR	.163	.658	.045	.247	.807	.820	1.220
AC_EN	-.838	.464	-.318	-1.807	.086	.882	1.134
AC_CH	-1.317	.559	-.437	-2.356	.029	.794	1.260
SS_EN	-.021	.089	-.054	-.231	.820	.496	2.016
SS_CH	-.097	.086	-.238	-1.135	.270	.620	1.613

Results of the multiple linear regression analysis showed that the linear regression model was moderately satisfactory, the goodness of fit reaching 45.5%. Results of T-tests for the regression coefficients showed that only AC CH (t=-2.356, p=0.029<0.05) affected F_AR significantly. The impact of AC CH on F_AR was significantly negative with the coefficient -1.317 < 0, indicating that F_AR was larger when AC CH was smaller. Besides, other variables like LA EN, LA CH, LR, AC EN, SS EN and SS CH did not affect the dependent variable F_AR significantly.

5.1.1.3 Influence on phonation time ratio

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in PTR (phonation time ratio) as an indicator of the development of speed fluency under conditions of low and high input rate, as PTR was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on S_PTR (change in phonation time ratio under low input rate conditions) was not significant. Results of the analysis with F_PTR (change in phonation time ratio under high input rate conditions) as the dependent variable are as follows.

Table 5.4 Coefficients of the regression model for F_PTR

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
(Constant)	.075	.051		1.489	.152		
LA_EN	.255	.210	.358	1.217	.238	.294	3.405
LA_CH	-.317	.203	-.489	-1.561	.134	.260	3.852
LR	-.166	.122	-.239	-1.358	.190	.820	1.220
AC_EN	-.009	.086	-.018	-.107	.916	.882	1.134
AC_CH	.362	.104	.624	3.485	.002	.794	1.260
SS_EN	.025	.017	.340	1.500	.149	.496	2.016
SS_CH	.036	.016	.458	2.261	.035	.620	1.613

Results of the multiple linear regression analysis showed that the linear regression model was moderately satisfactory and the goodness of fit reached 49.1%. Results of T-tests for the regression coefficients showed that AC CH (t=3.485, p=0.002<0.01) and SS CH (t=2.261, p=0.035<0.05) affected F_PTR significantly. The impact of AC CH on F_PTR was significantly positive, with the coefficient 0.362 > 0, indicating that F_PTR was larger when AC CH was larger. The impact of SS CH on F_PTR was significantly positive, with the coefficient 0.036 > 0, indicating that F_PTR was larger when SS CH was larger. Besides, other variables like LA EN, LA CH, LR, AC EN and SS EN did not affect the dependent variable F_PTR significantly.

5.1.1.4 Influence on mean length of run

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in MLR (mean length of run) as an indicator of the development of speed fluency under conditions of low and high input rate, as MLR was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on S_MLR (change in mean length of run under low input rate conditions) was not significant. Results of the analysis with F_MLR (change in mean length of run under high input rate conditions) as the dependent variable are as follows.

Table 5.5 Coefficients of the regression model for F_MLR

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
(Constant)	2.125	.813		2.613	.017		
LA_EN	4.437	3.370	.376	1.317	.203	.294	3.405
LA_CH	-7.401	3.268	-.688	-2.265	.035	.260	3.852
LR	-3.510	1.967	-.305	-1.785	.090	.820	1.220
AC_EN	-1.635	1.387	-.194	-1.179	.252	.882	1.134
AC_CH	4.479	1.671	.466	2.680	.014	.794	1.260
SS_EN	.294	.268	.242	1.099	.285	.496	2.016
SS_CH	.707	.257	.542	2.755	.012	.620	1.613

Results of the multiple linear regression analysis showed that the linear regression model was moderately satisfactory, and the goodness of fit reached 52.1%. Results of T-tests for the regression coefficients showed that LA CH (t=-2.265, p=0.035<0.05), AC CH (t=2.680, p=0.014<0.05) and SS CH (t=2.755, p=0.012<0.05) affected F_MLR significantly. The impact of LA CH on F_MLR was significantly negative, with the coefficient -7.401 < 0, indicating that F_MLR was larger when LA CH was smaller. The impact of AC CH on F_MLR was significantly positive, with the coefficient 4.479 > 0, indicating that F_MLR was larger when AC CH was larger. The impact of SS CH on F_MLR was significantly positive, with the coefficient 0.707 > 0, indicating that F_MLR was larger when SS CH was larger. Besides, other variables like LA EN, LR, AC EN and SS EN did not affect the dependent variable F_MLR significantly.

5.1.1.5 Diagnosis of the regression models

The above multiple linear regression models are diagnosed in the following in order to verify the accuracy of models.

a) Multicollinearity diagnosis

A multicollinearity problem may exist if the correlation of predictor variables is very high, with values of *r* above 0.80 or 0.90, and variance inflation factor (VIF) and *tolerance* can indicate whether a predictor has a strong linear relationship with another predictor(s) (Field, 2018). The VIF is "an index of the amount that the variance of each regression coefficient is increased relative to a situation in which all of the predictor variables are uncorrelated" and *tolerance* is its reciprocal (1/VIF) (Cohen et al., 2003, p. 423). A commonly used rule of thumb is that a VIF larger than 10 or a *tolerance* below 0.1 indicates a serious multicollinearity problem in the regression (ibid.). A more rigorous rule is that a VIF higher than 5 or a *tolerance* below 0.2 indicates a potential problem (Menard, 2002). The VIF values of the independent variables of the above regression models were all smaller than 5, indicating that the correlations between them were comparatively weak. Multicollinearity did not exist, and the above results of the regression models are reliable.

b) Normal test of residuals

The multiple linear regression model requires that residuals must obey normal distribution. "Multivariate normality is the assumption that each variable and all linear combinations of the variables are normally distributed", and for the multiple regression, residuals can be "screened for normality through the expected normal probability plot and the detrended normal probability plot" (Tabachnick & Fidell, 2012, p. 78,81). To examine the normality of data, the current research used the P-P plot, "a probability plot for assessing how closely two data sets agree, which plots the two cumulative distribution functions against each other" and the normality of data is assumed if the plot looks like a straight line or there is no curve (Das & Imon, 2016, p. 7). As shown in Figure 5.1, most plots are scattered near the diagonals. This indicates that residuals of the linear regression models obey normal distribution, which further verifies the accuracy of the results of the linear regression models.

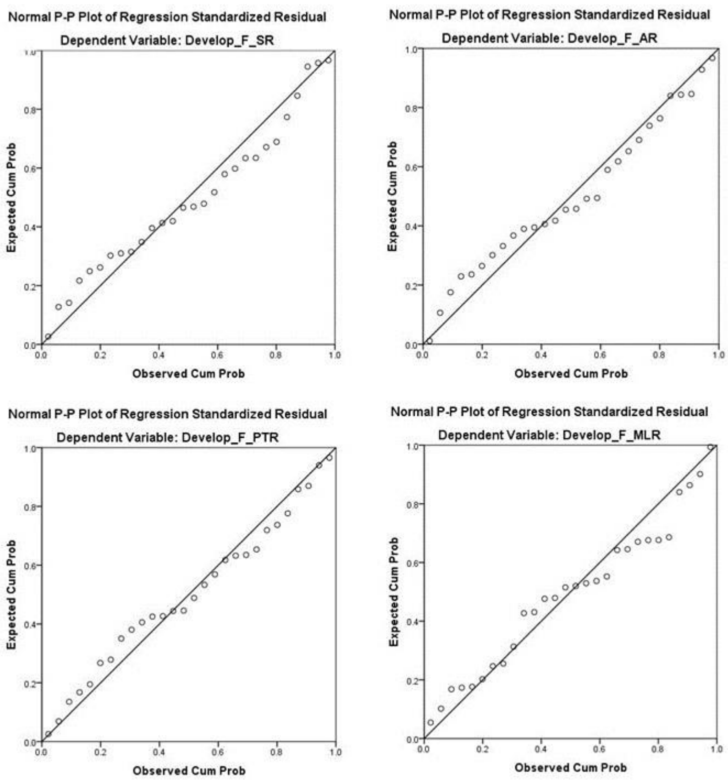


Figure 5.1 Normal P-P plots of models for speed fluency indicators

c) Heteroscedasticity diagnosis

Heteroscedasticity is the failure of homoscedasticity and is "caused either by the nonnormality of one of the variables or by the fact that one variable is related to some transformation of the other" (Tabachnick & Fidell, 2012, p. 85). Examination of a residual's scatterplot can offer tests of assumptions of linearity, normality, and homoscedasticity (ibid., p. 125). Results of the heteroscedasticity diagnosis are shown in Figure 5.2. The values of standardised residuals are all very small and mostly fall between -2 and 2. The residuals are distributed randomly, which means there is no heteroscedasticity for the models and further verifies the accuracy of the results of the linear regression models.

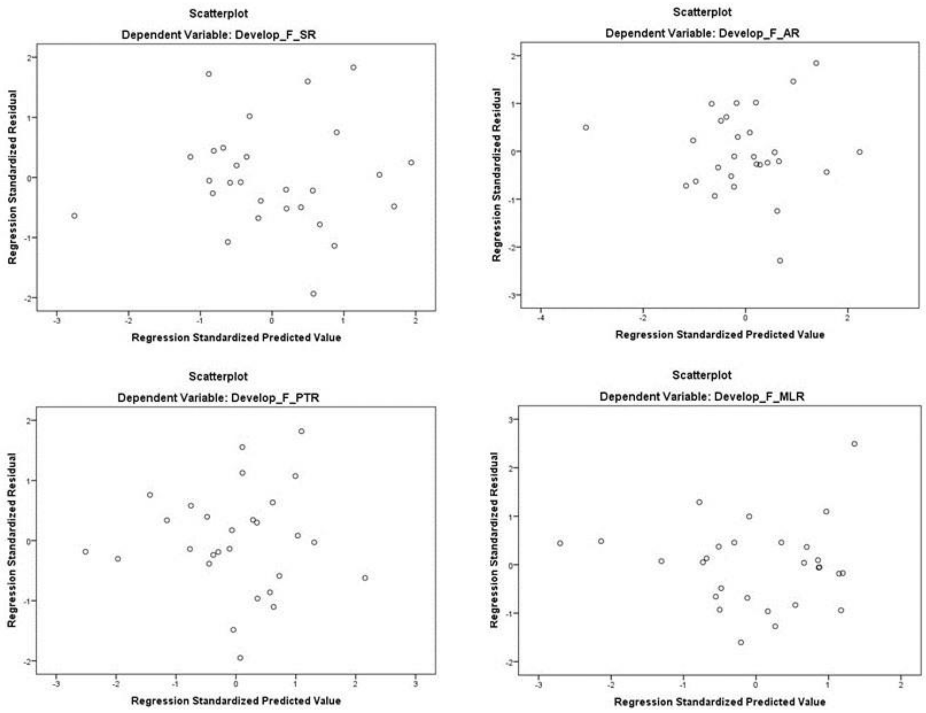


Figure 5.2 Scatterplots of models for speed fluency indicators

5.1.2 Influence of cognitive fluency on breakdown fluency development

For the development of breakdown fluency in the SI output, results of the multiple linear regression analyses revealed that the cognitive fluency measures had a significant impact on the change in SP mean and FP mean under conditions of low input rate and had significant impact on the change in SP length under conditions of high input rate. Besides, the cognitive fluency measures did not affect SP length or FP length significantly under low input rate conditions and they did not affect SP mean, FP mean or FP length significantly under high input rate conditions. Results of the analyses with significant results are presented in the following.

5.1.2.1 Influence on SP mean

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in SP mean (mean number of silent pauses per minute) as an indicator of the development of breakdown fluency under conditions of low and high input rate, as SP mean was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on F_SP mean (change in mean number of silent pauses per minute under high input rate conditions) was not significant. Results of the analysis with S_SP mean (change in mean number of silent pauses per minute under low input rate conditions) as the dependent variable are as follows.

Table 5.6 Coefficients of the regression model for S_SP mean

Model	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
(Constant)	-5.247	2.013		-2.606	.017		
LA_EN	19.840	8.343	.744	2.378	.027	.294	3.405
LA_CH	-14.968	8.091	-.616	-1.850	.079	.260	3.852
LR	14.488	4.870	.557	2.975	.007	.820	1.220
AC_EN	.717	3.433	.038	.209	.837	.882	1.134
AC_CH	3.234	4.138	.149	.782	.444	.794	1.260
SS_EN	.185	.662	.067	.280	.783	.496	2.016
SS_CH	-1.213	.635	-.411	-1.909	.071	.620	1.613

Results of the linear regression analysis showed that the linear regression model was moderately satisfactory, the goodness of fit reaching 42.5%. Results of T-tests for the regression coefficients showed that LA EN (t=2.378, p=0.027<0.05) and LR (t=2.975, p=0.007<0.01) affected S_SP mean significantly. The impact of LA EN on S_SP mean was significantly positive with the coefficient 19.840 > 0, indicating that S_SP mean was larger when LA EN was larger. The impact of LR on S_SP mean was significantly positive with the coefficient 14.488 > 0, indicating that S_SP mean was larger when LR was larger. Besides, other variables like LA CH, AC EN, AC CH, SS EN and SS CH did not affect the dependent variable S_SP mean significantly.

5.1.2.2 Influence on FP mean

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in FP mean as an indicator of the development of breakdown fluency under conditions of low and high input rate, as FP mean was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on F_FP mean (change in mean number of filled pauses per minute under high input rate conditions) was not significant. Results of the analysis with S_FP mean (change in mean number of filled pauses per minute under low input rate conditions) as the dependent variable are as follows.

Table 5.7 Coefficients of the regression model for S_FP mean

Model	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	2.100	1.254		1.675	.109		
LA_EN	-2.830	5.195	-.158	-.545	.592	.294	3.405
LA_CH	-2.495	5.038	-.153	-.495	.626	.260	3.852
LR	-1.855	3.032	-.106	-.612	.548	.820	1.220
AC_EN	-1.234	2.138	-.097	-.577	.570	.882	1.134
AC_CH	8.540	2.576	.585	3.315	.003	.794	1.260
SS_EN	-.094	.412	-.051	-.227	.823	.496	2.016
SS_CH	.538	.396	.272	1.360	.189	.620	1.613

Results of the linear regression analysis showed that the linear regression model was quite satisfactory, the goodness of fit reaching 50.5%. Results of T-tests for the regression coefficients showed that only AC CH (t=3.315, p=0.003<0.01) affected S_FP mean significantly. The impact of AC CH on S_FP mean was significantly positive with the coefficient 8.540 > 0, indicating that S_FP mean was larger when AC CH was larger. Besides, other variables like LA EN, LA CH, LR, AC EN, SS EN and SS CH did not affect the dependent variable S_FP mean significantly.

5.1.2.3 Influence on SP length

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in SP length as an indicator of the development of breakdown fluency under conditions of low and high input rate, as SP length was quantitative. Results of the analysis indicate that the impact of the cognitive fluency measures on S_SP length (mean duration of silent pauses under low input rate conditions) was not significant. Results of the analysis with F_SP length (mean duration of silent pauses under high input rate conditions) as the dependent variable are as follows.

Table 5.8 Coefficients of the regression model for F_SP length

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	.140	.160		.875	.392		
LA_EN	-.844	.665	-.430	-1.270	.219	.294	3.405
LA_CH	.510	.645	.285	.791	.438	.260	3.852
LR	.205	.388	.107	.528	.603	.820	1.220
AC_EN	.007	.274	.005	.027	.979	.882	1.134
AC_CH	-.798	.330	-.498	-2.420	.025	.794	1.260
SS_EN	-.076	.053	-.377	-1.447	.163	.496	2.016
SS_CH	-.040	.051	-.184	-.791	.438	.620	1.613

Results of the linear regression analysis showed that the linear regression model was acceptable, the goodness of fit reaching 32.6%. Results of T-tests for the regression coefficients showed that only AC CH (t=-2.420, p=0.025<0.05) affected F_SP length significantly. The impact of AC CH on F_SP length was significantly negative with the coefficient -0.798 < 0, indicating that F_SP length was larger when AC CH was smaller. Besides, other variables like LA EN, LA CH, LR, AC EN, SS EN and SS CH did not affect the dependent variable F_SP length significantly.

5.1.2.4 Diagnosis of the regression models

The above multiple linear regression models are diagnosed in the following in order to verify the accuracy of models.

a) Multicollinearity diagnosis

The VIF values of the independent variables of the regression models were all smaller than 5, indicating that the correlations between them were very weak. Multicollinearity did not exist, and the above results of the regression models are reliable.

b) Normal test of residuals

The multiple linear regression model requires that residuals must obey normal distribution. PP diagram was selected to investigate the normality of residuals. As shown in Figure 5.3, most plots are scattered near the diagonals. This indicates that residuals of the linear regression models obey normal distribution, which further verifies the accuracy of the results of the linear regression models.

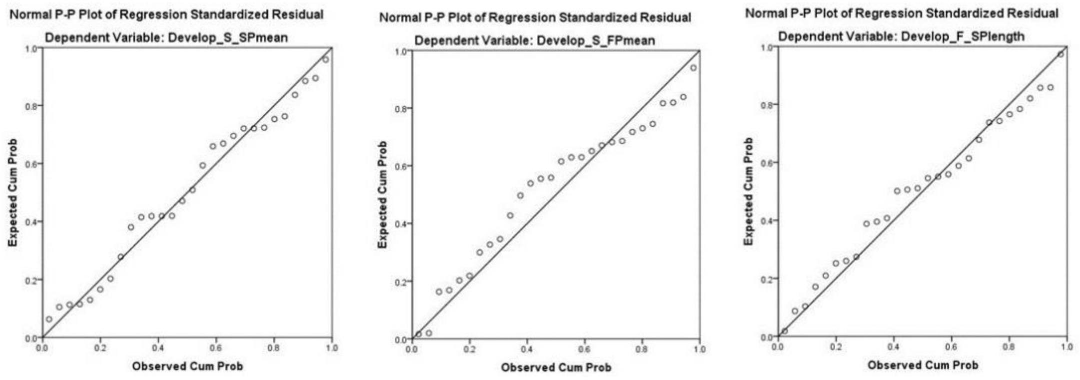


Figure 5.3 Normal P-P plots of models for breakdown fluency indicators

c) Heteroscedasticity diagnosis

Results of the heteroscedasticity diagnosis are shown in Figure 5.4, the values of standardized residuals are all very small and mostly fell between -2 and 2. Residuals are distributed comparatively randomly, which means there was no heteroscedasticity for the models and further verifies the accuracy of the results of the linear regression models.

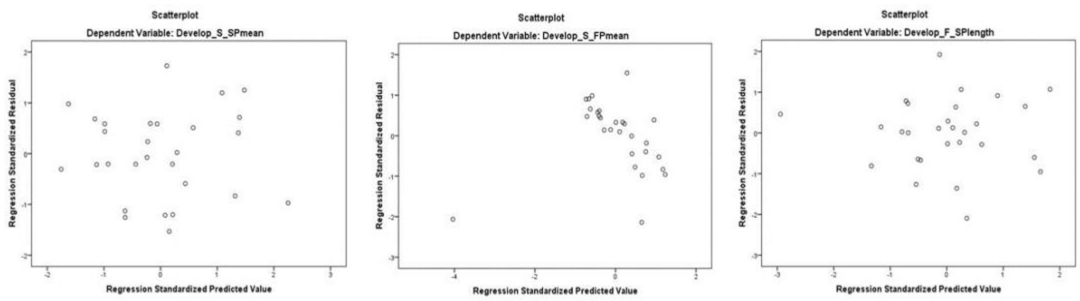


Figure 5.4 Scatterplots of models for breakdown fluency indicators

5.1.3 Impact of cognitive fluency on repair fluency development

For the development of repair fluency indicators in the SI output, results of the multiple linear regression analyses revealed that the cognitive fluency measures had a significant impact on the change in REPAIR length under conditions of high input rate. Besides, the cognitive fluency measures did not affect REPAIR mean significantly under low or high input rate conditions and they did not affect REPAIR length under high input rate conditions. Results of the analyses with significant results are presented in the following.

Multiple linear regression analysis was conducted to explore the impact of cognitive fluency on the change in REPAIR length as an indicator of the development of repair fluency under conditions of low and high input rate, as REPAIR length was quantitative. Results of the analysis with F_REPAIR length (mean duration of the sum of repairs, repetitions and false starts under high input rate conditions) as the dependent variable are as follows.

Table 5.9 Coefficients of the regression model for F_REPAIR length

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	-.574	.418		-1.374	.185		
LA_EN	-2.497	1.733	-.449	-1.441	.165	.294	3.405
LA_CH	4.857	1.680	.958	2.891	.009	.260	3.852
LR	-.147	1.011	-.027	-.145	.886	.820	1.220
AC_EN	-.563	.713	-.142	-.789	.439	.882	1.134
AC_CH	1.163	.859	.256	1.354	.191	.794	1.260
SS_EN	-.232	.138	-.405	-1.689	.107	.496	2.016
SS_CH	.170	.132	.277	1.292	.211	.620	1.613

Results of the linear regression analysis showed that the linear regression model was moderately satisfactory, the goodness of fit reaching 43.0%. Results of T-tests for the regression coefficients showed that LA CH (t=2.891, p=0.009<0.01) affected REPAIR length significantly. The impact of LA CH on F_REPAIR length was significantly positive with the coefficient 4.857 > 0, indicating that the F_REPAIR length was larger when the LA_CH was larger. Besides, other variables like LA EN, LR, AC EN, AC CH, SS EN and SS CH did not affect the dependent variable F_REPAIR length significantly.

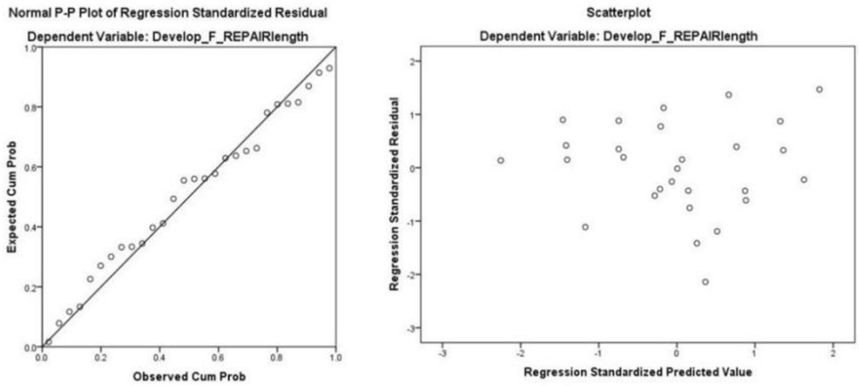


Figure 5.5 The P-P plot and scatterplot of models for F_REPAIR length

In order to verify the accuracy of the model, the normality of residuals and the heteroscedasticity of the above multiple linear regression model are diagnosed. As listed in Table 5.9, the VIF values of the independent variables of the regression model were all smaller than 2, indicating that the correlations between them were very weak. Multicollinearity did not exist and the above results of the regression model are reliable. As shown in the PP plot of Figure 5.5, most plots are scattered around the diagonal. It indicates that residuals of the linear regression model obey normal distribution, which further verifies the accuracy of the results of this linear regression model. The scatterplot shows that values of standardized residuals are all very small and mostly fall between -2 and 2. Residuals are distributed randomly, which indicates there is no heteroscedasticity for the model and further verifies the accuracy of the results of this linear regression model.

5.2 Summary and discussion of results

5.2.1 Summary of regression analyses results

The results of multiple linear regression analyses are summarised in Table 5.10, marking the explanatory power of the models (R^2) and the type of impact of the independent variables (positive or negative) that had a significant impact on each dependent variable (DV).

Table 5.10 Summary of results of standard multiple linear regression analyses

	Low Input Rate				**High Input Rate**			
	DV	**R^2**	**Predictors***		**DV**	**R^2**	**Predictors***	
Speed Fluency					SR	0.481	LA EN	(+)
							LA CH	(-)
					AR	0.455	AC CH	(-)
					PTR	0.491	AC CH	(+)
							SS CH	(+)
					MLR	0.521	LA CH	(-)
							AC CH	(+)
							SS CH	(+)
Breakdown Fluency	SP mean	0.425	LA EN	(+)	SP length	0.326	AC CH	(-)
			LR	(+)				
	FP mean	0.505	AC CH	(+)				
Repair fluency					REPAIR length	0.430	LA CH	(+)

Notes: * significant independent variables in correspondent regression models;
(+) positive impact; (-) negative impact.

As summarised in the above table, cognitive fluency measures could explain a large extent of the variance in the development of utterance fluency in trainee interpreters' SI output over an SI training period of one academic term. Cognitive fluency measures were significantly related to the development of breakdown fluency measures in the trainee interpreters' output under conditions of low input rate, and the regression models explained 42.5% and 50.5% of the variance of the change in mean number of silent pauses and mean number of filled pauses respectively. Cognitive fluency measures did not affect the mean duration of silent or filled pauses. Moreover, the explored cognitive fluency measures did not have a significant impact on the development of measures of speed fluency (SR, AR, PTR or MLR) and repair fluency (REPAIR mean or REPAIR length) under low input rate conditions. Under high input rate conditions, cognitive fluency measures had a significant impact on gains in all three dimensions of utterance fluency in trainee interpreters' SI output. The regression models explain around 50% of the variance of speed fluency development (with SR, AR, PTR and MLR as indicators separately), 32.6% of the variance of the change in mean duration of silent pauses and 43.0% of the variance of the change in mean duration of REPAIRs (repairs, repetitions and false starts). Moreover, the explored cognitive fluency measures did not affect the mean number of silent pauses, filled pauses and REPAIRs or the mean duration of filled pauses significantly under high input rate conditions.

The results of multiple linear regression analyses indicate that the impact of cognitive fluency measures on the development of SI utterance fluency was evidently stronger under high input rate conditions than under low input rate conditions. Cognitive fluency measures had a significant impact on changes in six indicators of three dimensions of utterance fluency under conditions of high input rate, while the impact was only significant with two indicators of breakdown fluency under low input rate conditions. This is in line with the fact that simultaneous interpreters are under higher cognitive load when the input rate is higher. With a low input rate, interpreters are under lower cognitive pressure, where the measures of cognitive fluency do not have so much impact on the development of utterance fluency in SI.

Moreover, the results showed that most of the cognitive fluency measures that had a significant impact on the development of utterance fluency were measures of the Chinese versions of the cognitive tasks. Since the language direction of the SI tasks was English to Chinese (L2-L1), it implies that the efficiency of cognitive processes involved in the target language production stage has a main significant impact on the development of utterance fluency, rather than the efficiency of processes involved in the source language comprehension stage. It is worth further exploration in future research to see whether this is true in the other language direction (L1-L2).

5.2.2 Discussion of the role of individual cognitive fluency measures

For further discussion of the results of the multiple regression analyses, it is worth restating the values of independent and dependent variables adopted in this research: the CV (coefficient of variance) of the measures of cognitive fluency as independent variables and the development of utterance fluency, i.e. changes in indicators of utterance fluency (measures in the post-test minus their counterparts in the pre-test), as the individual dependent variables. A smaller CV represents more efficient processing of the cognitive processes. Results regarding the impacts of individual cognitive fluency measures are summarised in Table 5.11.

Table 5.11 Summary of the influence of individual cognitive fluency measures

Predictors *	**Low Input Rate**		**High Input Rate**	
	DV	Influence	DV	Influence
LA EN	SP mean	(+)	SR	(+)
LA CH			SR	(-)
			MLR	(-)
			REPAIR length	(+)
LR	SP mean	(+)		
AC CH	FP mean	(+)	AR	(-)
			PTR	(+)
			MLR	(+)
			SP length	(-)
SS CH			PTR	(+)
			MLR	(+)

Notes: * significant predictors of cognitive fluency measures;
(+) positive impact; (-) negative impact

5.2.2.1 Influence of lexical access and retrieval

The efficiency of lexical access of the source and target languages had a different impact on the development of utterance fluency in SI. Results of analyses showed that the efficiency of English lexical access (LA EN) had a significantly positive impact on the change in SP mean under low input rate conditions and in SR under high input rate conditions. It implied that a lower efficiency (bigger CV) of lexical access in the source language contributed to a bigger change in these two fluency indicators under low and high input rate conditions respectively. The efficiency of lexical access to the target language (LA CH) did not have a significant impact on the development of utterance fluency under conditions of low input rate, but under high input rate conditions, LA CH had a significantly negative impact on the development of SR and MLR and a significantly positive impact on REPAIR length. It implied that higher efficiency (smaller CV) of lexical access in the target language contributed significantly to more gains in SR and MLR in the interpreting output and fewer gains in the length of REPAIRs under conditions of high input rate.

The efficiency of lexical retrieval only contributed slightly when the cognitive load was comparatively low and did not affect the development of utterance fluency significantly under conditions of high cognitive load. The efficiency of lexical retrieval as measured in the word translation task, which involved the conversion of words from source to target language and required verbal articulation, only correlated significantly with the change in SP mean under conditions of low input rate. Empirical research exploring the role of lexical retrieval in interpreting performance is scarce (Cai et al., 2015; Christoffels et al., 2003). The study conducted by Christoffels et al. (2003) found that the retrieval of translation equivalence was an important contributor to SI performance. However, Cai et al. (2015) failed to find a meaningful contribution of lexical retrieval efficiency to consecutive interpreting performance. Differences in experiment design, participant profiles and modes of interpreting make comparison difficult: the word translation task chosen by Cai et al. (2015) was the translation recognition task, whereas Christoffels et al. (2003) and the current research used the translation production task, which not only tapped into the retrieval of translation equivalents but also their verbal production; participants in Christoffels et al.'s (2003) study were untrained bilinguals with relatively high L2 proficiency, while those in Cai et al.'s (2015) study and the current research were unbalanced bilinguals who were trainee interpreters at the initial stage of interpreting training; Cai et al. (2015) focused on the performance of consecutive interpreting, whereas Christoffels et al. (2003) and the current study explored SI performance. Despite these differences, the current research offers a different view of the role of lexical retrieval in the SI performance of trainee interpreters and finds that the impact of lexical retrieval efficiency on SI fluency development differs under different conditions of cognitive load. It would be valuable for future studies to explore how lexical retrieval affects the performance of interpreters of different levels of expertise.

The results imply that the roles of lexical access and lexical retrieval processes in utterance fluency development in SI are divergent: the efficiency of lexical access evidently had a strong impact on utterance fluency development under higher cognitive load conditions, though it only affected the change in SP mean under conditions of lower cognitive load; the efficiency of lexical retrieval only contributed minimally to the change in SP mean when the cognitive load was comparatively low. Lexical access involves the recognition of linguistic forms and arriving at the correct lexical entry from the mental lexicon, and lexical retrieval involves the selection of lexical concepts that are subsequently transformed into linguistic forms (Levelt, 1989; Snellings et al., 2002). It has been stated that many components of lexical access and lexical retrieval are interchangeable, though the order in which the process components are activated is reversed (Levelt, Roelofs, & Meyer, 1999; Snellings et al., 2002). The results of the current research emphasise the importance of differentiating lexical access and lexical retrieval when exploring their role in SI performance.

5.2.2.2 Influence of linguistic attention control

The efficiency of linguistic attention control in the source language (AC EN) did not affect the development of utterance fluency under either low or high input rate conditions, while the efficiency of linguistic attention control in the target language (AC CH) affected gains in multiple indicators of utterance fluency of the SI output.

This further provides evidence for the view that the efficiency of target language processing seems to play a more important role in SI.

Filled pauses are interruptions of speech flow and reflect hesitations and uncertainties in the information formulation stage and brief attention to planning or retrieval. Under conditions of low cognitive load, AC CH only had a significantly positive impact on the change in FP mean, implying that more efficient shifting (smaller CV) indicates a smaller change in the mean number of filled pauses. Under conditions of high cognitive load, the impact was significant for different dimensions of utterance fluency: AC CH had a significantly negative impact on the development of AR and SP length and a significantly positive impact on the development of PTR and MLR. This implied that higher efficiency (smaller CV) of linguistic attention control in the target language (AC CH) contributed to more gains in AR and the mean duration of silent pauses under conditions of high input rate, while higher efficiency (smaller CV) of linguistic attention control in the target language contributes to a smaller change in PTR and MLR. To summarise, the results indicate that more efficient shifting brought more AR improvement, accompanied by longer duration of silent pauses between utterances and less efficient shifting led to more gains in phonation time ratio and longer runs of utterances when the cognitive load was comparatively high.

The observed patterns of the negative impact of AC CH on AR and SP length and the positive impact of AC CH on PTR and MLR under high input rate conditions were consistent. AR did not compound measures of pauses as it was calculated as the number of characters in SI output (including disfluencies) divided by total speaking time apart from pauses. As shown above, higher efficiency of linguistic attention control (smaller

CV) contributes to an increase in AR. More AR gains mean that trainee interpreters execute the articulation of the interpreted message faster and spare more time for other cognitive processes like source language comprehension, making longer silent pauses possible. Silent pauses in SI indicate more focused attention on comprehension or cognitive coordination. Faster AR also indicates that trainee interpreters would manage to process information and articulate more efficiently, leading to smaller phonation time ratio and shorter mean length of run.

The shifting tested in the current research was language-directed attention control, differing from previous studies in which shifting was as a function of general-domain cognitive control. Though the isolated nature of previous related studies makes it difficult to compare, the findings of the current research generally support previous studies in that shifting, as a function of cognitive control, is important in interpreting performance (Babcock & Vallesi, 2017; Timarová et al., 2014). Previous studies (e.g. Dong & Liu, 2016) indicated that interpreting experience contributed to a cognitive advantage in shifting efficiency while the current research provides evidence that shifting efficiency contributes to the development of interpreting performance. It is worth further exploration of whether domain-general and domain-specific cognitive efficiency make similar contributions to interpreting performance. A direction for future related work could be to explore different cognitive tasks and more functions of cognitive control.

5.2.2.3 Influence of working memory capacity

Working memory capacity in the target language (SS CH), elicited from the speaking span task, had a significant impact on utterance fluency development in L2-L1 SI performance under conditions of high input rate, but the impact was not significant under conditions of low input rate. Working memory capacity in the source language (SS EN) did not affect the development of utterance fluency significantly. This conforms with the domain-specific view of working memory in the sense that different cognitive resources are required for different domains of processing (Miyake, 2001; Shah & Miyake, 1996), though it is inconsistent with previous findings (Cai et al., 2015; Christoffels et al., 2003). Cai et al. (2015) observed that L2 working memory spans had a stronger relationship with L2-L1 consecutive interpreting performance compared with L1 working memory spans. Christoffels et al. (2003) also found that working memory capacity in L2 might be more important in SI than the L1 capacity. Differences in the tasks testing working memory span, interpreting modes, participant profiles and explored aspects of interpreting performance might explain the discrepancies.

Different span tasks test different aspects of memory and involve different stages of language processing, which might lead to different findings. Cai et al. (2015) tested both L1 (Chinese) and L2 (English) listening span and speaking span. Listening span measured in L2 (but not in L1) significantly correlated with consecutive interpreting performance. L2 speaking span significantly correlated with interpreting performance at both time slots, whereas L1 speaking span only correlated with interpreting performance in the pretest. The working memory task chosen by Christoffels et al. (2003) was a reading span task, while the current study administered a speaking span task in L1 and L2 versions to pay more attention to the production stage.

Different modes of interpreting place inherently different demands on interpreters' cognitive resources. Comprehension of the source language is relatively more capacity-demanding than production in consecutive interpreting while the simultaneity of comprehension and production in SI makes both processes cognitively highly demanding. This partly explains why L2 working memory span had a stronger relationship with consecutive interpreting performance in Cai et al.'s (2015) study. Apart from the differences in span tasks, the participants in the study conducted by Christoffels et al. (2003) were untrained bilinguals, while Cai et al. (2015) and the currents research both used student interpreters as participants. A period of interpreting training might develop an interpreter advantage that bilinguals do not possess, which might contribute to the discrepancies.

In terms of the explored aspects of interpreting performance, both Cai et al. (2015) and Christoffels et al. (2003) asked two judges to rate the interpreting performance. The criteria in the study by Cai et al. (2015) were information (accuracy and completeness) and target language use (grammar and appropriateness). Two measures were used by Christoffels et al. (2003), i.e. how well the content of ten sentences was translated and how well the texts were interpreted for the whole recording. In both previous studies, the content was taken into consideration, which differed from the fluency construct explored in the current research. In addition, the data in the current analyses were obtained through objective measurement of indicators of fluency instead of human rating. It is possible that L1 and L2 working memory resources play different roles in different constructs of interpreting performance. The construct of fluency should be distinguished from overall interpreting performance and fidelity and should be investigated separately.

Moreover, working memory capacity only correlated with some indicators of the development of SI utterance fluency under conditions of high cognitive load. Speaking span in the target language (SS CH) had a significantly positive impact on gains in PTR and MLR under conditions of high cognitive load. Larger memory span allows more information to be stored, which makes a higher phonation time ratio and longer utterance runs possible. This conforms to the proposition that working memory effects are manifest in capacity-demanding tasks (Just & Carpenter, 1992; Shah & Miyake, 1996) and provides further evidence of the role of working memory in SI performance.

Working memory is regarded as an important internal cognitive factor affecting both L1 and L2 language processing, but the issue of whether memory plays an independent role in language production has not reached a consistent conclusion. Hierarchical regression analyses were conducted to further investigate if SS CH made an independent contribution to gains in PTR and MLR as indicators of SI utterance fluency development under conditions of high cognitive load. With other cognitive fluency measures controlled in the first layer and SS CH in the second layer, the results of hierarchical regression analyses showed that L1 working memory capacity (SS CH) made an independent contribution to the change in both PTR ($\triangle R_2 = 0.13$, $\triangle F = 5.113$, $Sig.\triangle F = 0.035 < 0.05$) and MLR ($\triangle R_2 = 0.182$, $\triangle F = 7.593$, $Sig.\triangle F = 0.012 < 0.05$). This finding provides evidence of the unique contribution of L1 working memory to the development of SI speed fluency.

5.2.3 Added explanatory power of lexical retrieval and working memory capacity

Processes involved in cognitive fluency include “the speed and efficiency of semantic retrieval, the handling of the attention-focusing demands inherent in utterance construction, operations in working memory, among others" (Segalowitz, 2016, p. 82). Processing speed and stability of lexical access and linguistic attention control are important aspects of cognitive fluency. Processing stability is needed in order to escape pressures on cognitive resources. Linguistic attention control is required in order to focus and refocus processing resources in changing situations (Segalowitz, 2010). It is important to realise that cognitive fluency includes more than the above-mentioned aspects, which have been dealt with most regarding L2 learning. It has been noted that cognitive fluency also includes working memory operations and conceptualising skills (De Jong, 2013; Segalowitz, 2016), but these aspects are seldom explored in past research.

The current research explores the role of cognitive fluency in utterance fluency of SI performance, which opens new perspectives for cognitive fluency research. Cognitive fluency explored in the current research involves both L1 and L2 efficiency of mobilising and integrating underlying cognitive processes, which both play an important part in SI. The current research follows previous experimental designs in studies of L2 fluency to examine the efficiency of lexical access and linguistic attention control. Meanwhile, features of SI should be considered and accommodated as SI is a much more complex language processing task than second language production. It not only requires simultaneous source language comprehension and target language production but also involves language conversion and coordination of memory resources. Considering the unique features of SI and based on previous studies which have provided evidence for the link between SI performance and lexical retrieval and working memory capacity (as reviewed in 2.2.1), this research attempts to tap into more aspects of cognitive fluency and adds a lexical retrieval task and a speaking span task in its exploration into the role of cognitive fluency in SI utterance fluency.

The explanatory power of regression models, with or without measures of lexical retrieval and working memory capacity, for SI utterance development are compared below to verify the additional contribution of these two tasks. The independent variables in Model 1 are measures of lexical access and linguistic attention control tasks in both L1 and L2 versions, including LA EN, LA CH, AC EN and AC CH. The independent variables in Model 2 are measures of four cognitive tasks, i.e. lexical access, linguistic attention control and working memory capacity tasks in L1 and L2 versions and a lexical retrieval (L2-L1) task, including LA EN, LA CH, AC EN, AC CH, LR, SS EN and SS CH. The lexical retrieval task (word translation) is from L2 to L1 (English-Chinse), which is consistent with the language direction of the SI tasks. The speaking span task adopts the composite strict span value and all other tasks use the CV value as indexes for the efficiency of cognitive processes.

Table 5.12 Comparison of regression models with and without measures of lexical retrieval and working memory capacity

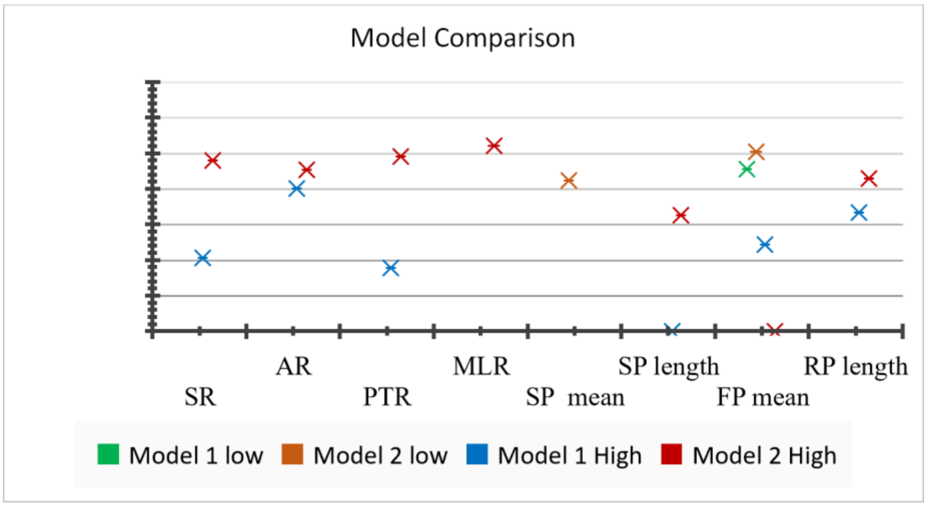
	DV	Model 1		Model 2		
		R^2	Predictor*	R^2	Predictor*	
Low Input Rate	SP mean			0.425	LA EN LR	↑
	FP mean	0.456	AC CH	0.505	AC CH	↗
High Input Rate	SR	0.206	LA EN	0.481	LA EN LA CH	↗↗
	AR	0.401	AC CH	0.455	AC CH	↗
	PTR	0.179	AC CH	0.491	AC CH SS CH	↗↗
	MLR			0.521	LA CH AC CH SS CH	↑
	SP length			0.326	AC CH	↑
	FP mean	0.245	AC CH			↓
	REPAIR length	0.335	LA CH	0.430	LA CH	↗

Notes: * significant independent variables in correspondent regression models;
↑ significantly affected by the independent variable(s) of Model 2, but not Model 1;
↓ significantly affected by the independent variable(s) of Model 1, but not Model 2;
↗ slight increase of explanatory power of Model 2, compared to Model 1;
↗↗ remarkable increase of explanatory power of Model 2, compared to Model 1.

The above table illustrates comparisons of Model 1 and Model 2, listing dependent variables (indicators of SI utterance fluency) that are significantly affected, the correspondent explanatory power of the model (R^2) and significant independent variables (measures of cognitive fluency). The comparison generally shows an evidently stronger explanatory power of Model 2, which has added cognitive predictors of lexical retrieval and working memory span.

As shown in Table 5.12, under conditions of low input rate, independent variables of Model 1 (only lexical access and linguistic attention control measures as predictors) did not affect SP mean significantly, while SP mean was significantly affected by LA EN and LR in Model 2, which explained 42.5% of the variance of the SP mean. For the FP mean, the explanatory power of Model 2 was only slightly higher (50.50% vs. 45.60%). Under conditions of high input rate, the explanatory power of Model 2 was remarkably higher than that of Model 1 for SR (48.10% vs. 20.60%) and PTR (49.10% vs. 17.90%). For AR and REPAIR length, Model 2 explained slightly more variance of the dependent variables than Model 1. The predictors of Model 1 did not have a significant impact on MLR and SP length, while Model 2 explained 52.10% and 32.60% of their variance respectively. Finally, for FP mean, Model 1 explained 24.50% of the variance, while the predictors of Model 2 had no significantly effect.

The following graph visualises comparisons of the explanatory power of Model 1 and Model 2. Comparisons of Model 1 and Model 2 under conditions of low and high input rate respectively show that independent variables of Model 2 generally showed a higher explanatory power.



Notes: Low refers to conditions of low input rate; High refers to conditions of high input rate;
Model 1 refers to the regression model with measures of lexical access and linguistic attention control as predictors;
Model 2 refers to the regression model with measures of lexical access, lexical retrieval, linguistic attention control and working memory capacity as predictors.

Chart 5.1 Comparison of the explanatory power of Model 1 and Model 2

In order to verify whether the added measures of lexical retrieval and working memory capacity make an independent contribution in explaining the variance of utterance fluency development, hierarchical regression analyses were conducted with each dependent variable (utterance fluency indicators are listed in Table 5.12). With measures of lexical access and linguistic attention control (LA EN, LA CH, AC EN, AC CH) in the first layer, lexical retrieval indexes (LR) in the second layer and working memory span indexes (SS EN, SS CH) in the third layer, results of the hierarchical regression analyses showed that LR significantly increased the explanatory power of the original model (with LA and AC measures as predictors) for SP mean under low input conditions ($\triangle R_2 = 0.162$, $\triangle F = 5.174$, $Sig.\triangle F = 0.033 < 0.05$). Working memory capacity (SS EN, SS CH) significantly increased the explanatory power of the original models (with LA, AC and LR measures as predictors) for SR ($\triangle R^2 = 0.273$, $\triangle F = 5.260$, $Sig.\triangle F = 0.015 < 0.05$), PTR ($\triangle R^2 = 0.311$, $\triangle F = 6.108$, $Sig.\triangle F = 0.009 < 0.05$) and MLR ($\triangle R^2 = 0.326$, $\triangle F = 6.794$, $Sig.\triangle F = 0.006 < 0.05$) under high input rate conditions. For the other dependent variables, the additional contribution of lexical retrieval and working memory capacity measures to the overall explanatory power of the regression models did not reach significance, though the overall model power was increased.

In a nutshell, measures of lexical retrieval and working memory generally increased the explanatory power of cognitive fluency measures for the utterance fluency development of trainee interpreters' SI output. Lexical retrieval only made an independent contribution to the SP mean in the SI output under conditions of low cognitive load. Working memory capacity made an independent contribution to gains in speech rate, phonation time ratio and mean length of run as indicators of trainee interpreters' SI utterance fluency under conditions of high cognitive load, but its role was not significant under conditions of low cognitive load. Moreover, the composite strict speaking span adopted in this research took into account processing accuracy, processing efficiency and storage ability, which was further proved to be an effective index exploring working memory capacity. The findings provide evidence for the effectiveness of including lexical retrieval and working memory capacity in the exploration of cognitive fluency, which has implications for the theoretical framework of cognitive fluency in studies of both interpreting and bilingual language production.

Chapter 6 The effects of cognitive fluency, SI training and input rate on SI utterance fluency

This chapter focuses on the influence of three factors, i.e. cognitive fluency, SI training and cognitive load, on utterance fluency in SI output. Repeated measures ANOVA analyses were conducted to explore the effects of cognitive fluency, SI training and input rate on utterance fluency indicators of the trainee interpreters' SI output and to investigate whether there was interaction between the three factors. The ANOVA analyses were conducted with training and input rate as within-subjects variables and each of the cognitive fluency measures as the between-subjects variable. Cognitive fluency measures were recoded into a dichotomous variable based on the mean value. The mean difference was significant at the 0.05 level and the *Bonferroni* method was used for confidence interval adjustment for multiple comparisons.

6.1 Data analysis

6.1.1 Effects of cognitive fluency, training and input rate on speed fluency in SI

A series of repeated measures ANOVA analyses found that indicators of speed fluency in SI output were significantly affected by one or more measures of cognitive fluency, SI training and/or input rate of source speeches. Their effects on SR, AR, MLR and PTR are discussed in detail in this section.

6.1.1.1 Effects of cognitive fluency, training and input rate on SR

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on SR (speech rate) and the results are discussed in this section.

In terms of the effects of cognitive fluency, results of the analyses showed that the effect of LR on SR was significant [$F(1,26)=4.697, p=0.040$] and the effects of other cognitive fluency measures on SR did not reach significance. The pairwise comparison showed that the SR of high LR efficiency group was higher than the low LR efficiency group by 0.294 and the difference was significant, $p<0.05$. Results of the main effects of training and input rate on SR are summarized in Table 6.1.

Table 6.1 Summary of results for the effects of training and input rate on SR

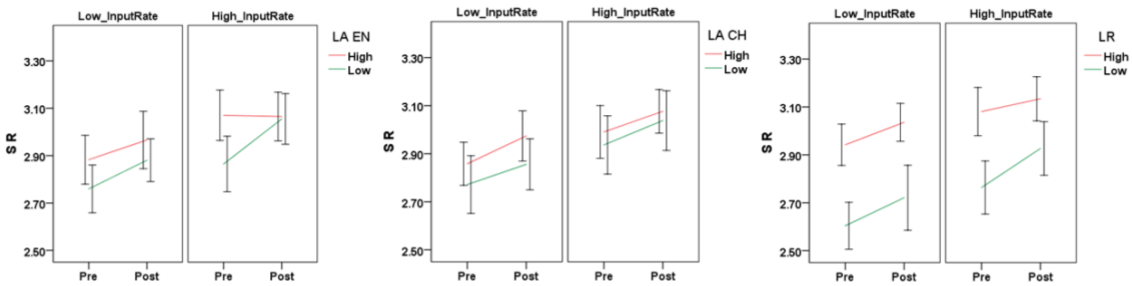
	Between-subject Variables	Main Effects	
		Training	**Input Rate**
	LA EN	** [F(1,26)=8.266, p=.008]	** [F(1,26)=19.748, p=.000]
	LA CH	* [F(1,26)=7.184, p=.013]	** [F(1,26)=21.072, p=.000]
SR	LR*	** [F(1,26)=8.469, p=.007]	** [F(1,26)=21.328, p=.000]
	AC EN	* [F(1,26)=7.337, p=.012]	** [F(1,26)=23.257, p=.000]
	AC CH	* [F(1,26)=7.385, p=.012]	** [F(1,26)=19.550, p=.000]
	SS EN	** [F(1,26)=9.584, p=.005]	** [F(1,26)=28.428, p=.000]
	SS CH	** [F(1,26)=7.843, p=.009]	** [F(1,26)=20.397, p=.000]

Note: * $p<0.05$; ** $p<0.01$

As shown in Table 6.1, the effects of training on SR were all significant, $p<0.05$, which indicates that there was a significant change in the SR between the pre-test and post-test SI output. Pairwise comparisons showed that the mean SR in the pre-test was lower than that in the post-test and the difference was statistically significant. The effects of input rate on SR were all significant, $p < 0.01$, which indicates that there was a significant change in the SR under conditions of low and high input rate. Pairwise comparisons showed that the mean SR under low input rate conditions was lower than that under high input rate conditions and the difference was statistically significant.

Besides, there was significant interaction effects between the input rate and SS EN [$F(1,26)=6.990$, $p=0.014 < 0.05$] and other interaction effects among cognitive fluency measures, SI training and input rate did not reach significance. Analyses of simple effects showed that the effect of SS EN was significant under conditions of high input rate, [$F(1.26)=4.328$, $p=0.047 < 0.05$], with the SR for the low SS EN group lower than that for the high SS EN group by 0.140; the effect of SS EN was not significant under conditions of low input rate [$F(1.26)=1.017$, $p=0.323 > 0.05$]; the effect of input rate was significant for the high SS EN group [$F(1.26)=27.830$, $p < 0.001$] and the effect of input rate approached significance for the low SS EN group, [$F(1.26)=4.214$, $p=0.0503$], with the SR under high input rate conditions faster than that under low input rate conditions for the high and low SS EN groups by 0.228 and 0.077 respectively. Therefore, the effect of SS EN was only significant under conditions of high input rate and the effect of input rate was more evident for the high SS EN group.

The following section discusses the effects of cognitive fluency measures, training and input rate on SR based on the visualised charts. The changes in SR in the SI output are discussed for participants in the high and low efficiency groups of cognitive fluency under conditions of high and low input rate in the pre-test and post-test. A smaller CV represents more efficient cognitive processing. The high cognitive fluency group (visualised by the red line) represents the group with higher efficiency of correspondent cognitive process (low CV index) and the low cognitive fluency group (visualised by the green line) represents the group with low efficiency of cognitive process (high CV index).



Notes: —— High for high efficiency of the cognitive process;

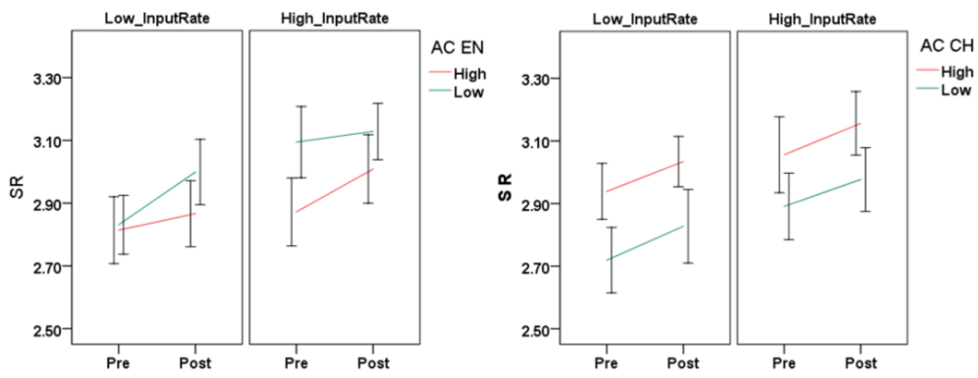
—— Low for low efficiency of the cognitive process.

Chart 6.1 Influence of LA, LR, training and input rate on SR

In terms of the efficiency of LA EN, LA CH and LR, the SR of the high efficiency groups was generally higher than that of the low efficiency groups, though only the difference between the low and high LR groups reached significance. The mean SR in the post-test was overall faster than that in the pre-test for both low and high efficiency groups under conditions of both low and high input rate, except for the LA EN group under conditions of high input rate, where the SR in the post-test performance was similar to that of the pre-test.

The SR of the high LA EN group did not change after training in the high cognitive load context, and the pattern was different from the low input rate context. It is noteworthy that under high input rate conditions the SR of the high LA EN group was much faster than that of the low LA EN group in the pre-test while the SR of the low LA EN group grew to a similar level with that of the high LA EN group in the post-test.

This is probably an indication of the ceiling effect since the SR of the high LA EN group in the pre-test was already quite high.

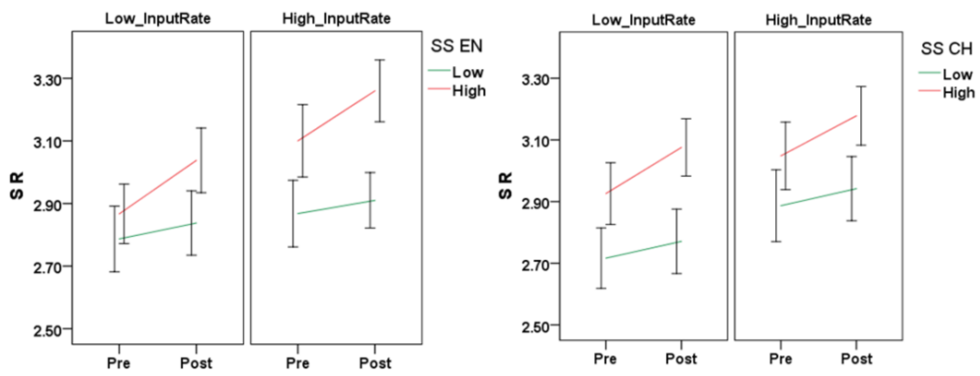


Notes: —— High for high efficiency of the cognitive process;

—— Low for low efficiency of the cognitive process.

Chart 6.2 Influence of AC, training and input rate on SR

As for the efficiency of the flexibility of attention control, the influence of AC EN and AC CH exhibited different patterns. The SR of the high efficiency AC EN group was lower than that of the low AC EN group while the SR of the high AC CH group was higher than that of the low AC CH group, though the difference between the two groups did not reach significance. This indicates a different influence of L1 (CH) and L2 (EN) on the SR of the interpreted output. The mean SR in the post-test was higher than that in the pre-test for both low and high AC EN and AC CH groups under conditions of both low and high input rate. The SR under high input rate conditions was faster than that under low input rate conditions for both the high and low AC EN and AC CH groups in both pre-test and post-test.



Notes: —— High for a high capacity of working memory;

—— Low for a low capacity of working memory.

Chart 6.3 Influence of SS, training and input rate on SR

For working memory capacity, the influence of SS EN and SS CH exhibited similar trends. The SR of the high SS EN group was higher than that of the low SS EN group, though the difference between the two groups did not reach significance; the trend was similar for the SS CH groups. It indicates that a larger working memory capacity (speaking span) in either L1 or L2 might contribute to a higher SR in the interpreted output. Furthermore, the effect of SS EN was only significant under conditions of high input rate according to the analysis of the simple effects. The mean SR in the post-test was higher than that in the pre-test for both low and high SS EN and SS CH groups under conditions of both low and high input rate. The SR under high input rate conditions was higher than that under conditions of low input rate for both the high and low SS EN and SS CH groups in both pre-test and post-test.

6.1.1.2 Effects of cognitive fluency, training and input rate on AR

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on AR (articulation rate).

The effects of all cognitive fluency measures on AR were not significant. But there were significant two-way interaction effects between the input rate and LA EN [F(1,26)=4.800, p=0.038＜0.01], input rate and LA CH [F(1,26)=10.162, p=0.004＜0.01] and input rate and LR [F(1,26)=4.640, p=0.041＜0.01]. Other two-way interaction effects among cognitive fluency measures, SI training and input rate did not reach significance. Analyses of simple effects for two-way interactions showed that: The effects of input rates were significant for groups of high LA EN efficiency [F(1.26)=9.734, p=0.004＜0.05], low LA EN efficiency [F(1.26)=38.667, p＜0.001], high LA CH efficiency [F(1.26)=11.033, p=0.003＜0.01], low LA CH efficiency [F(1.26)=50.322, p＜0.001], high LR efficiency [F(1.26)=15.946, p＜0.001], and low LR efficiency [F(1.26)=32.07, p＜0.001] separately. Pairwise comparisons showed that the AR under conditions of high input rate was higher than that under low input rate conditions for all LA EN, LA CH and LR groups. The effects of LA EN, LA CH and LR were not significant under conditions of both high and low input rates.

Training and input rate were found to have significant impact on AR. Results of the main effects of training and input rate on AR are summarized in Table 6.2.

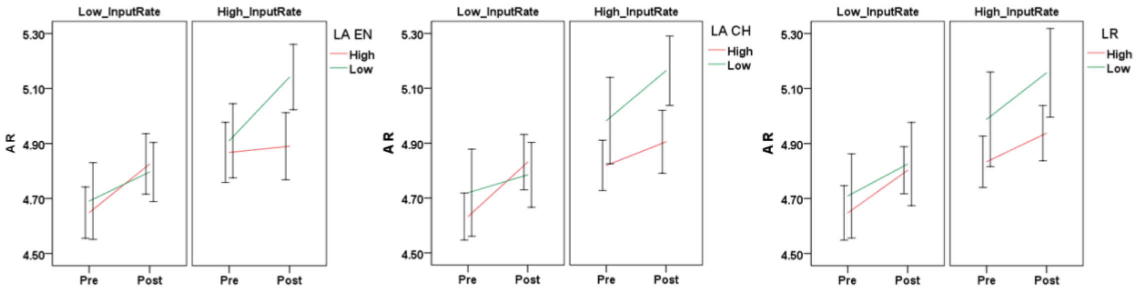
Table 6.2 Summary of results for the effects of training and input rate on AR

	Between-subject Variables	Main Effects	
		Training	Input Rate
AR	LA EN	* [F(1,26)=7.080, p=.013]	** [F(1,26)=43.601, p=.000]
	LA CH	* [F(1,26)=6.686, p=.016]	** [F(1,26)=56.805, p=.000]
	LR	* [F(1,26)=6.570, p=.017]	** [F(1,26)=47.982, p=.000]
	AC EN	* [F(1,26)=6.406, p=.018]	** [F(1,26)=38.573, p=.000]
	AC CH	** [F(1,26)=8.031, p=0.009]	** [F(1,26)=36.688, p=.000]
	SS EN	* [F(1,26)=6.767, p=.015]	** [F(1,26)=38.953, p=.000]
	SS CH	* [F(1,26)=7.231, p=.012]	** [F(1,26)=39.004, p=.000]

Note: * $p < 0.05$; ** $p < 0.01$; the effects of cognitive fluency are not significant.

As shown in Table 6.2, the main effects of training on AR were all significant, $p < 0.05$, which indicates that there was a significant change in the AR after training. The pairwise comparison showed that the mean AR in pre-test was lower than that in the post-test and the difference was statistically significant. The main effects of input rate on AR were all significant, $p < 0.01$, which indicates that there was a significant change in the AR under conditions of low and high input rate. The pairwise comparison showed that the mean AR under low input rate conditions was lower than that under high input rate conditions and the difference was statistically significant.

The following section illustrates the effects of cognitive fluency measures, training and input rate on AR based on the visualised charts, and the changes in AR in the SI output are discussed.



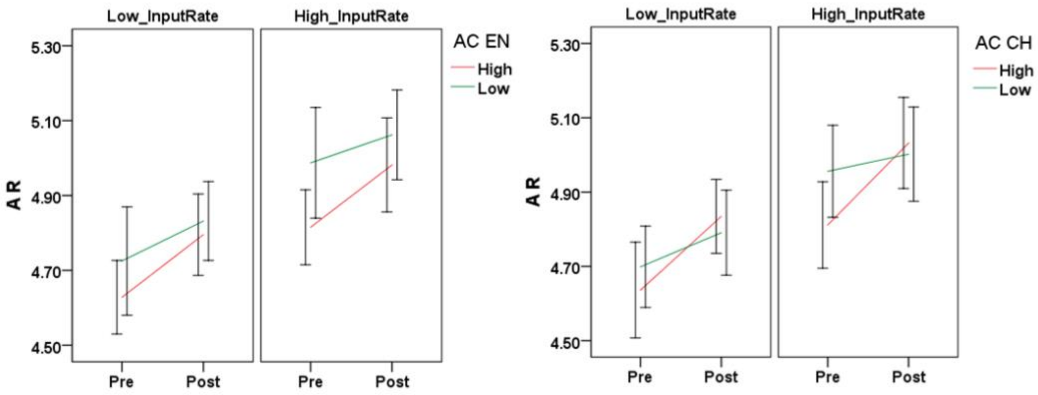
Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.4 Influence of LA, LR, training and input rate on AR

As illustrated in Chart 6.4, the AR of the high cognitive fluency group was higher than that of the low cognitive fluency group under conditions of high input rate in terms of LA EN, LA CH and LR efficiency, though the differences between the AR of the low and high LA EN, LA CH and LR efficiency groups did not reach significance. Moreover, the mean AR in the post-test was higher than that in the pre-test for both the low and high efficiency groups under conditions of both low and high input rate. The AR under high input rate conditions was higher than that under the low input rate conditions for both the high and low LA EN, LA CH and LR groups in both pre-test and post-test; the same pattern was observed for the low LA EN, LA CH and LR groups in the post-test, but not for the high LA EN, LA CH and LR groups, which saw similar AR under low and high input rates in the post-test.

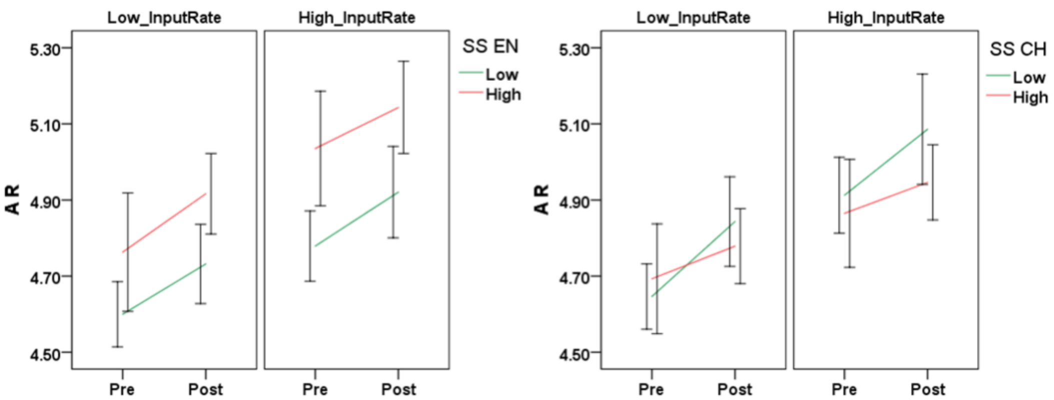
As for the efficiency of the flexibility of attention control, the AR of the high AC EN group was lower than that of the low AC EN group under both low and high input rate conditions in both pre-test and post-test, as illustrated in Chart 6.5. The AR of the high AC CH group was lower than that of the low AC CH group in the pre-test under both low and high input rates, while after training, the AR of the high AC CH group grew significantly and became slightly higher than that of the low AC CH group in the post-test.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.5 Influence of AC, SS, training and input rate on AR

Moreover, the mean AR in the post-test was faster than that in the pre-test for all AC groups under conditions of both low and high input rate. The AR under high input rate conditions was faster than that under low input rate conditions for all AC groups in both pre-test and post-test.



Notes: —— High represents a high capacity of working memory;
—— Low represents a low capacity of working memory.

Chart 6.6 Influence of SS, training and input rate on AR

For working memory capacity, SS EN and SS CH influenced AR differently, though the difference between the two groups did not reach significance. The AR of the high SS EN group was higher than that of the low SS EN group. However, the AR of the high SS CH group was overall lower than that of the low SS CH group, except under low input rate conditions in the pre-test where the high SS CH group had slightly higher AR than the low SS CH group. This indicates a different impact of L1 and L2 working memory capacity on AR in the interpreted output.

The mean AR in the post-test was faster than that in the pre-test for both low and high SS EN and SS CH groups under conditions of both low and high input rates. The AR under high input rate conditions was faster than that under low input rate conditions for both the high and low SS EN and SS CH groups in both pre-test and post-test.

6.1.1.3 Effects of cognitive fluency, training and input rate on MLR

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on MLR (mean length of run).

In terms of cognitive fluency, the effect of LR on MLR was significant [$F(1,26)=5.001$, $p=0.034$] and the effects of other cognitive fluency measures on MLR were not significant. The pairwise comparison showed that the MLR of the high LR efficiency group was longer than that of the low LR efficiency group by 0.758 and the difference was significant, $p=0.034<0.05$, which indicates that participants with higher efficiency of LR produced longer utterance runs.

Besides, there were significant interaction effects between the input rate and SS EN [$F(1,26)=4.229$, $p=0.0499$]. Other interaction effects among cognitive fluency, SI training and input rate did not reach significance. Analyses of the simple effects are showed below: the effects of input rates were significant for both the low SS EN group [$F(1.26)=9.385$, $p=0.005<0.01$] and the high SS EN group [$F(1.26)=28.873$, $p<0.001$]; The MLR under low input rate condition was shorter than that under high input rate conditions by 0.379 and 0.767 for the low and high SS EN groups respectively. The effects of SS EN were not significant for conditions of low or high input rates.

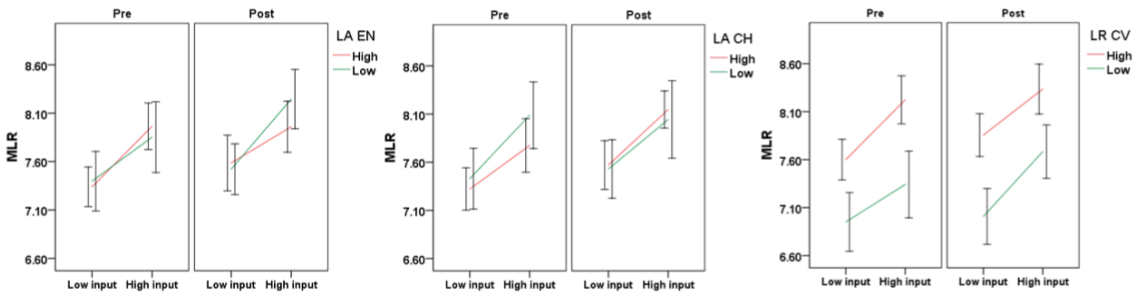
Table 6.3 Summary of results for the effects of input rate on MLR

	Between-subject Variables	**Main Effects (Input Rate)**
MLR	LA EN	** [F(1,26)=29.491, p=.000]
	LA CH	** [F(1,26)=29.359, p=.000]
	LR	** [F(1,26)=26.632, p=.000]
	AC EN	** [F(1,26)=30.483, p=.000]
	AC CH	** [F(1,26)=32.001, p=.000]
	SS EN	** [F(1,26)=36.813, p=.000]
	SS CH	** [F(1,26)=29.528, p=.000]

Note: ** $p<0.01$; the effects of training are not significant.

The effects of training on MLR did not reach significance, which indicates that there was no significant change in the MLR between the pre-test and post-test. The effects of input rate on MLR were all significant, $p<0.01$, as shown in Table 6.3. It indicates that there was a significant change in the MLR under conditions of low and high input rates. Pairwise comparisons showed that the mean MLR under conditions of low input rate was shorter than that under high input rate conditions and the difference was statistically significant.

The effects of cognitive fluency, training and input rate on MLR are illustrated in the following based on the visualised charts, and the changes in MLR in the SI output are discussed.

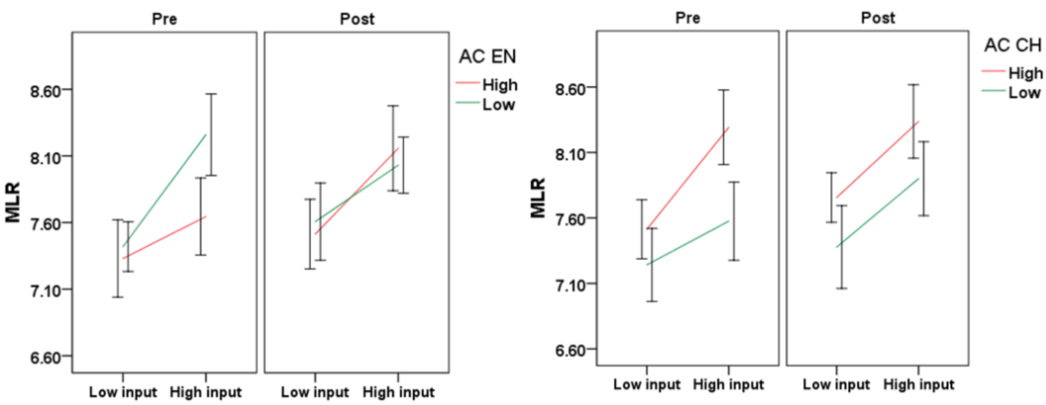


Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.7 Influence of LA, LR, training and input rate on MLR

LR efficiency had a significant impact on MLR. As illustrated in the above chart, the MLR of the high LR efficiency group was longer than that of the low LR efficiency group under conditions of both low and high input rates in both pre-test and post-test. The effects of LA EN and LA CH on MLR were not significant, and their patterns were irregular, as shown in the chart. The MLR under high input rate conditions was longer than that under low input rate conditions for all LA EN, LA CH and LR groups in both the pre-test and the post-test.

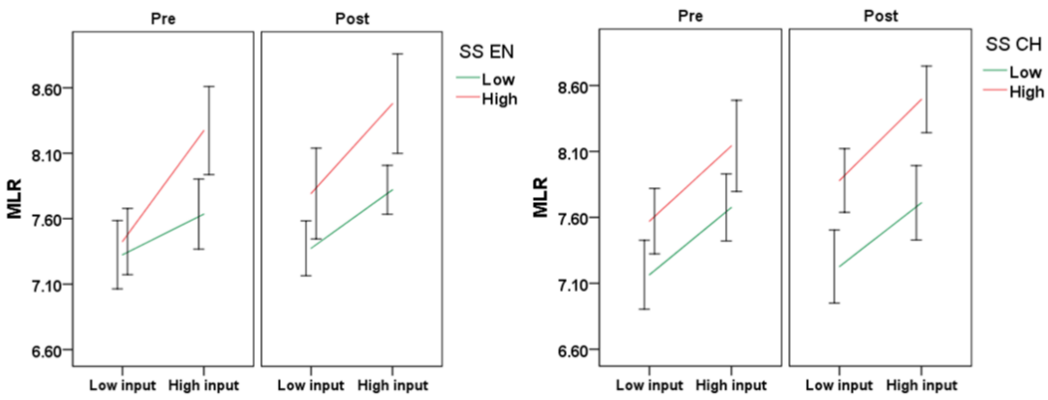


Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.8 Influence of AC, training and input rate on MLR

As for linguistic attention control, it is shown that the MLR of the high AC CH group was longer than that of the low AC CH group under both low and high input rate conditions in both the pre-test and the post-test, though AC measures did not have a significant impact on MLR. For AC EN, the MLR of the high AC EN group was shorter than that of the low AC EN group in the pre-test under conditions of both low and high input rates, while in the post-test, the high AC EN group produced slightly shorter MLR under low input rate conditions and longer MLR under high input rate conditions than the low AC EN group.

There was no significant difference between the mean MLR in the pre-test and the post-test. The MLR under high input rate conditions was significantly longer than that under the low input rate conditions for all AC groups in both the pre-test and the post-test.



Notes: —— High represents a high capacity of working memory;
—— Low represents a low capacity of working memory.

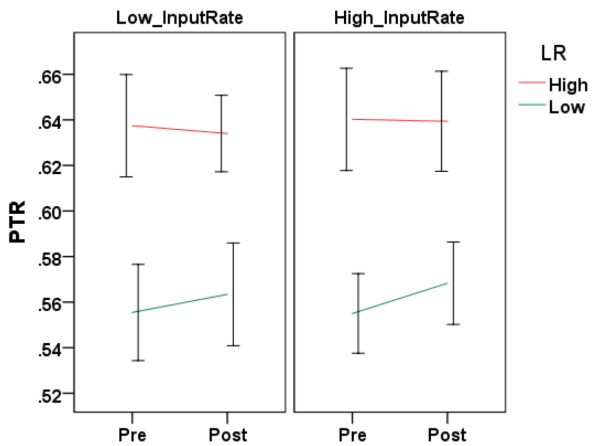
Chart 6.9 Influence of SS, training and input rate on MLR

In terms of working memory capacity, the MLR of the high SS EN group was longer than that of the low SS EN group, and the SS CH groups shared the same pattern, though the difference between the two groups did not reach significance. This indicates that larger working memory capacity in both L1 and L2 contributed to longer MLR in the interpreted output. There was no significant difference between the MLR in the pre-test and the post-test. The MLR under high input rate conditions was significantly longer than that under low input rate conditions for all SS groups in both the pre-test and the post-test.

6.1.1.4 Effects of cognitive fluency, training and input rate on PTR

Repeated measures ANOVA analyses were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on PTR (phonation time ratio). The results show that the effect of LR on PTR was significant [$F(1,26) = 6.820$, $p = 0.015$] and the effects of all other cognitive fluency measures on PTR were insignificant. Neither SI training nor input rate had a significant impact on PTR. The pairwise comparison showed that the PTR of the high LR efficiency group was longer than that of the low LR efficiency group by 0.077.

As illustrated in Chart 6.10, the PTR of the high LR efficiency group was much higher than that of the low LR efficiency group under both low and high input rate conditions and in both the pre-test and the post-test. This indicates that participants with higher LR efficiency had a higher ratio of phonation time in their SI output. The differences in PTR between conditions of low and high input rates and between the pre-test and the post-test were minimal. There were no interaction effects between cognitive fluency measures, training and input rate.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.10 Influence of LR, training and input rate on PTR

6.1.2 Effects of cognitive fluency, training and input rate on breakdown fluency in SI

Results of repeated measures ANOVA analyses found that indicators of breakdown fluency in the SI output were significantly affected by one or more measures of cognitive fluency, SI training and/or input rate of source speeches. Their effects on SP mean, SP length, FP mean and FP length are discussed in detail in this section.

6.1.2.1 Effects of cognitive fluency, training and input rate on SP mean

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on SP mean (the mean number of silent pauses per minute). The effects of cognitive fluency measures and training on SP mean were not significant, which indicates that there was no significant change in the SP mean between low and high efficiency cognitive fluency groups and between the pre-test and post-test. Results of the main effects of input rate on SP mean are summarized in Table 6.4.

Table 6.4 Summary of results for the effects of input rate on SP mean

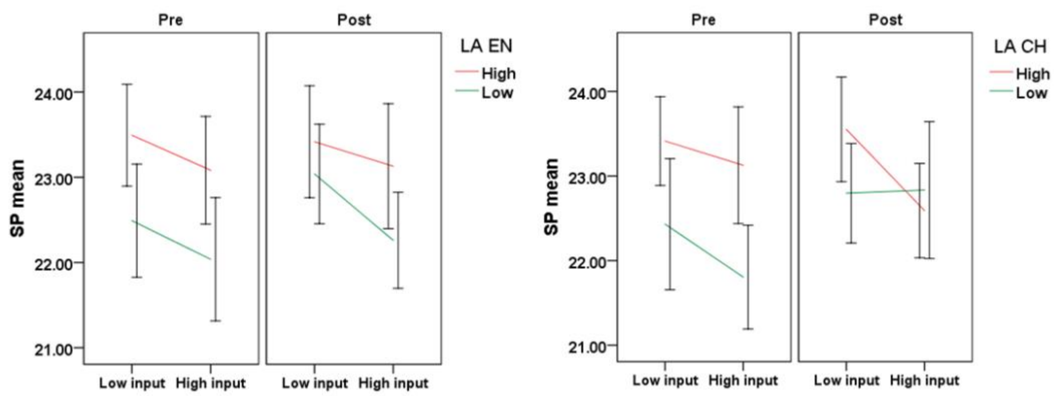
	Between-subject Variables	**Main Effects (Input Rate)**
SP mean	LA EN	* [F(1,26)=5.213, p=0.031]
	LA CH	* [F(1,26)=4.657, p=0.040]
	AC EN	* [F(1,26)=4.800, p=0.038]
	AC CH	* [F(1,26)=7.841, p=0.010]
	SS EN	* [F(1,26)=5.053, p=0.033]
	SS CH	* [F(1,26)=5.605, p=0.026]

Note: * $p < 0.05$; the effects of cognitive fluency and training are both insignificant.

The effects of input rate on SP mean were all significant, except when with LR efficiency as the between-subject variable, which indicates that there was a significant change in the SP mean under conditions of low and high input rates in most cases. Pairwise comparisons showed that the SP mean under low input rate conditions was in most cases larger than that of high input rate conditions and the difference was statistically significant.

Besides, there was a significant interaction effect between the input rate and AC CH [$F(1,26)=8.270$, $p=0.008<0.01$]. Other interaction effects among cognitive fluency measures, SI training and input rate did not reach significance. Analyses of simple effects showed that: The effects of input rates were significant for the high AC CH efficiency group [$F(1.26)=15.035$, $p=0.001<0.01$], but not for the low AC CH efficiency group [$F(1.26)=0.003$, $p=0.956>0.05$]. The SP mean under the low input rate conditions was larger than that under high input rate conditions by 1.055 for the high AC CH efficiency group. The effects of AC CH were not significant under conditions low or high input rates.

The effects of cognitive fluency, training and input rate on SP mean are illustrated below based on the visualised charts, and the changes in SP mean in the interpreted output are discussed.

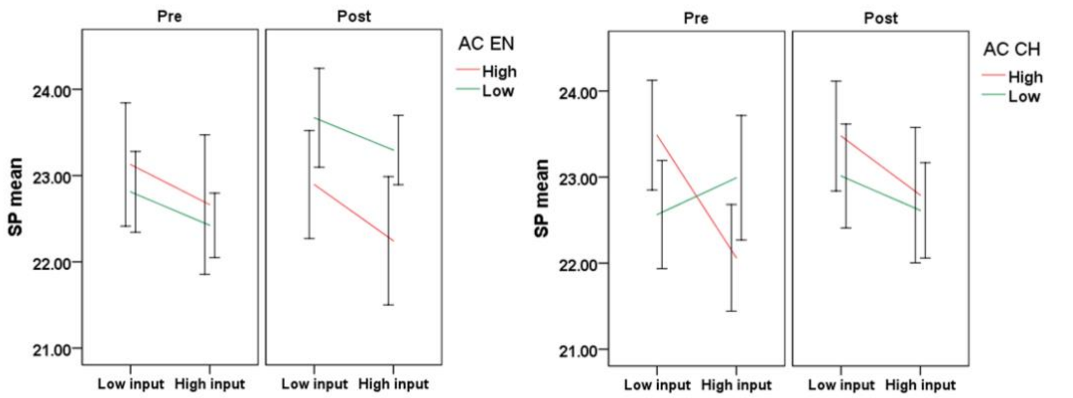


Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.11 Influence of LA, training and input rate on SP mean

As illustrated in Chart 6.11, the SP mean under high input rate conditions was significantly smaller than that under low input rate conditions for both high and low LA EN groups in both the pre-test and the post-test. The SP mean of the high LA EN group was larger than that of the low LA EN group under both low and high input rate conditions in both pre-test and post-test, though the impact of LA EN on SP mean did not reach significance.

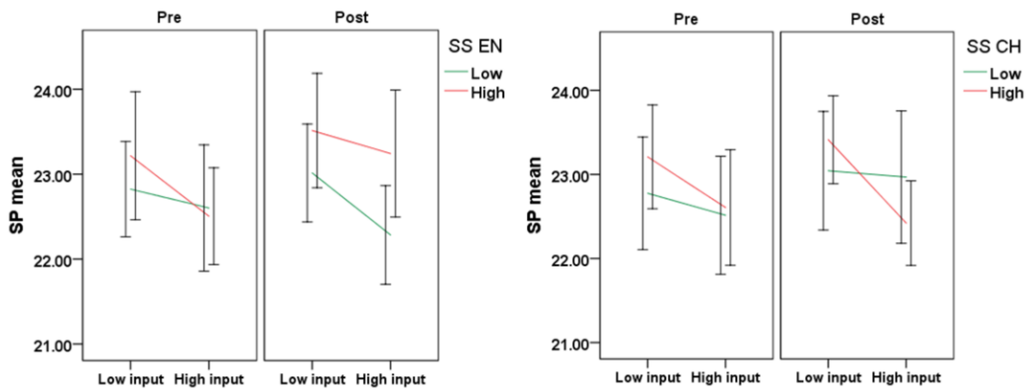
In terms of LA CH efficiency, the SP mean of the high LA CH group was larger than that of the low LA CH group under both low and high input rate conditions in the pre-test. In the post-test, however, the SP mean of the high LA CH group was decreased significantly under high input rate conditions while the SP mean was similar under low and high input rate conditions for the low LA CH group. This might indicate that the high LA CH group managed to decrease the number of SP under high cognitive load while the low LA CH group failed to do so.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.12 Influence of AC, training and input rate on SP mean

In terms of AC efficiency, the SP mean under high input rate conditions was significantly smaller than that under low input rate conditions for both the high and low AC EN groups in both the pre-test and the post-test, which pattern was the same for AC CH groups in the post-test. In the pre-test, however, the change in SP mean exhibited a different trend, with SP mean under high input rate conditions decreasing for the high AC CH group and increasing slightly for the low AC CH group compared to that under low input rate conditions.



Notes: High represents a high capacity of working memory;

Low represents a low capacity of working memory.

Chart 6.13 Influence of SS, training and input rate on SP mean

As illustrated in the above chart, the SP mean under high input rate conditions was significantly smaller than that under low input rate conditions for both low and high SS EN and SS CH groups in both pre-test and post-test.

6.1.2.2 Effects of cognitive fluency, training and input rate on SP length

Repeated measures ANOVA were conducted to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on SP length (mean duration of silent pauses).

In terms of cognitive fluency, the effect of LR on SP length approached significance [$F(1,26)=3.999$, $p=0.056$]. The pairwise comparison showed that the SP length of high LR efficiency group was shorter than that of the low LR efficiency group by 0.210, which implied that participants with a higher efficiency of LR paused for a shorter time while interpreting. The effects of other cognitive fluency measures on SP length were not significant. The effects of input rate on SP length did not reach significance either. Results of the main effects of training are summarized in Table 6.5. Besides, there were no significant interaction effects between cognitive fluency, SI training and input.

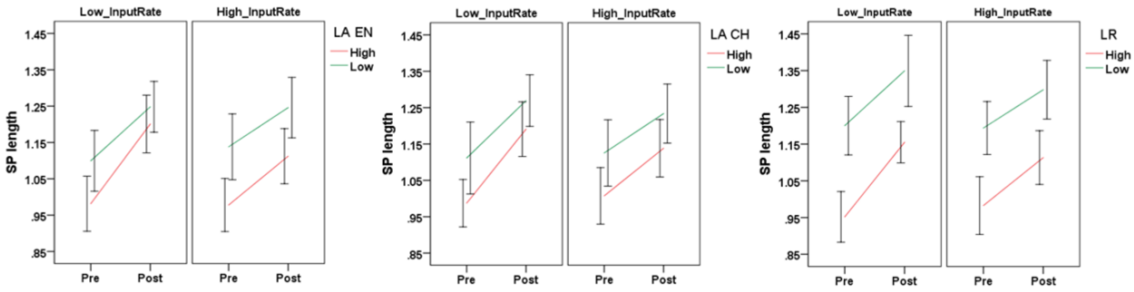
Table 6.5 Summary of results for the effects of training on SP length

	Between-subject Variables	**Main Effects (Training)**
SP length	LA EN	** [F(1,26)=34.798, p=.000]
	LA CH	** [F(1,26)=32.443, p=.000]
	LR	** [F(1,26)=29.234, p=.000]
	AC EN	** [F(1,26)=33.078, p=.000]
	AC CH	** [F(1,26)=34969, p=.000]
	SS EN	** [F(1,26)=33.808, p=.000]
	SS CH	** [F(1,26)=34.744, p=.000]

Note: ** $p < 0.01$; the effects of input rate are not significant.

The effects of training on SP length were all significant, $p < 0.001$, which indicates that there was a significant change in the SP length in the pre-test and the post-test. Pairwise comparisons showed that the mean SP length in pre-test was shorter than that in the post-test and the difference was statistically significant.

The effects of cognitive fluency, training and input rate on SP length are illustrated in the following based on the visualised charts and the changes in SP length in the interpreted output are discussed.

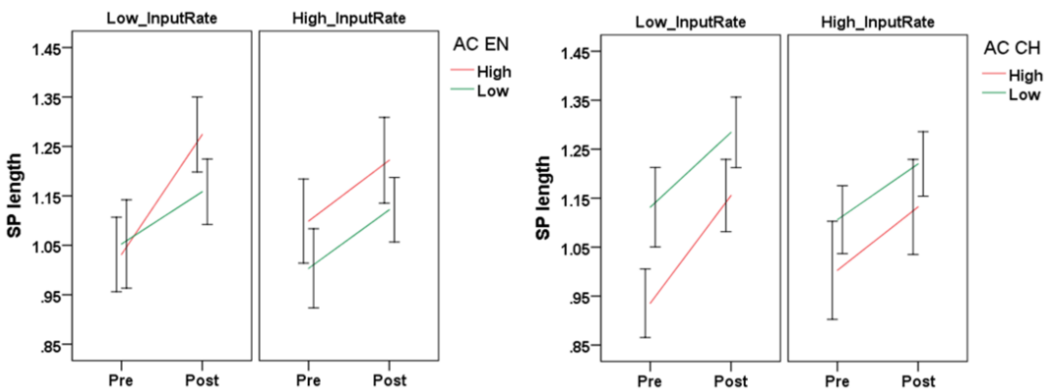


Notes: — High represents high efficiency of the cognitive process;

— Low represents low efficiency of the cognitive process.

Chart 6.14 Influence of LA, LR, training and input rate on SP length

As illustrated in Chart 6.14, the SP length of the high LA EN group was shorter than that of the low LA EN group under both low and high input rate conditions and in both the pre-test and the post-test, and the pattern was the same for the LA CH groups and the LR groups, though only the difference between the low and high LR groups approached significance. This implies that the mean duration of silent pauses was shorter in the SI output of trainee interpreters with higher LA and LR efficiency. The SP length in the post-test was longer than in the pre-test for each LA and LR group under both low and high input rate conditions, which implied that trainee interpreters paused for a longer duration between their interpreted utterances after training.

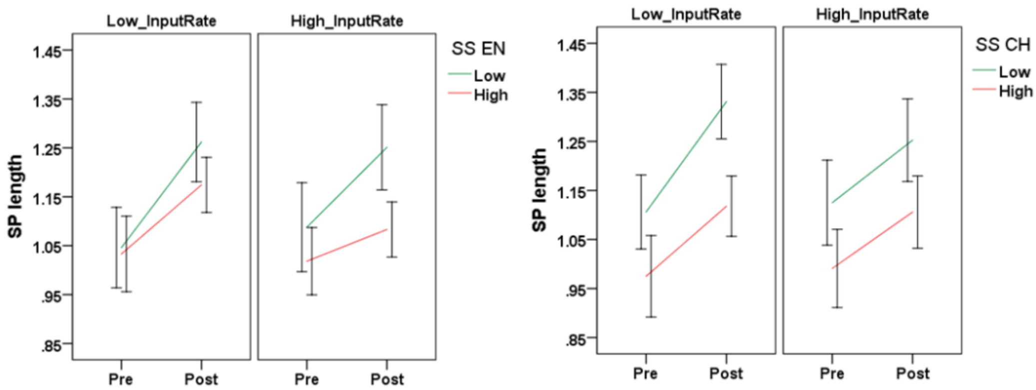


Notes: — High represents high efficiency of the cognitive process;

— Low represents low efficiency of the cognitive process.

Chart 6.15 Influence of AC, training and input rate on SP length

The SP length of the high AC EN group was generally longer than that of the low AC EN group except under low input rate conditions in the pre-test, where the SP length of high AC EN group was slightly shorter. The SP length of the high AC CH group was shorter than that of the low AC CH group under conditions of both low and high input rates in both the pre-test and the post-test, though the differences were not significant. Moreover, the SP length in the post-test was longer than in the pre-test for all AC groups under both low and high input rate conditions in the pre-test and the post-test.



Notes: —— High represents a high capacity of working memory;

—— Low represents a low capacity of working memory.

Chart 6.16 Influence of SS, training and input rate on SP length

The SP length of the high SS EN group was shorter than that of the low SS EN group under conditions of both low and high input rate in the pre-test and the post-test, and the pattern was the same for the SS CH groups, though the differences in SP length were not significant. This indicates that trainee interpreters with larger working memory capacity tended to have shorter silent pauses in their SI output. Moreover, the SP length in the post-test was longer than that in the pre-test for all SS groups under both low and high input rate conditions in both the pre-test and the post-test.

6.1.2.3 Effects of cognitive fluency, training and input rate on FP mean

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on FP mean (the mean number of filled pauses per minute).

The effects of cognitive fluency measures and training on FP mean were not significant, which indicates that there was no significant change in the FP mean between the low and high efficiency cognitive fluency groups and between the pre-test and the post-test. Results of the main effects of input rate on FP mean are summarized in the following table.

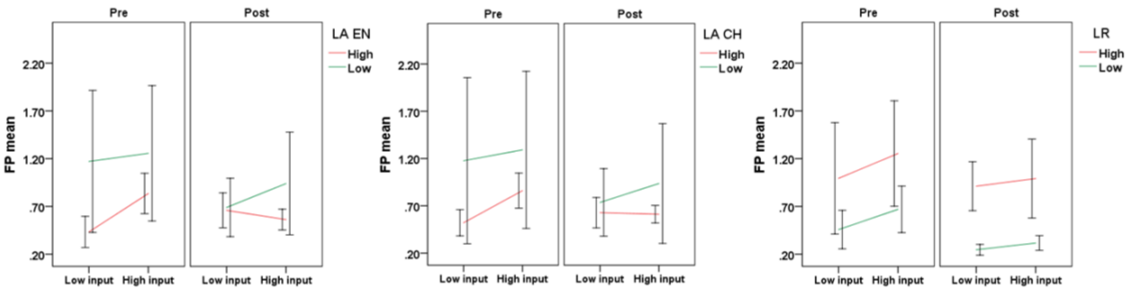
Table 6.6 Summary of results for the effects of input rate on FP mean

	Between-subject Variables	**Main Effects (Input Rate)**
FP mean	LA EN	* [F(1,26)=5.333, p=.029]
	LA CH	* [F(1,26)=5.202, p=.031]
	LR	* [F(1,26)=4.655, p=.040]
	AC EN	* [F(1,26)=6.289, p=.019]
	AC CH	* [F(1,26)=5.945, p=.022]
	SS EN	* [F(1,26)=6.508, p=.017]
	SS CH	* [F(1,26)=5.634, p=.025]

Note: * $p < 0.05$; the effects of training are not significant.

The effects of input rate on FP mean were all significant, which indicates that there was a significant change in the FP mean under conditions of low and high input rates. Pairwise comparisons showed that the FP mean under the low input rate conditions was smaller than that under high input rate conditions and the difference was statistically significant. It indicates more hesitations in the SI output under high input rate conditions.

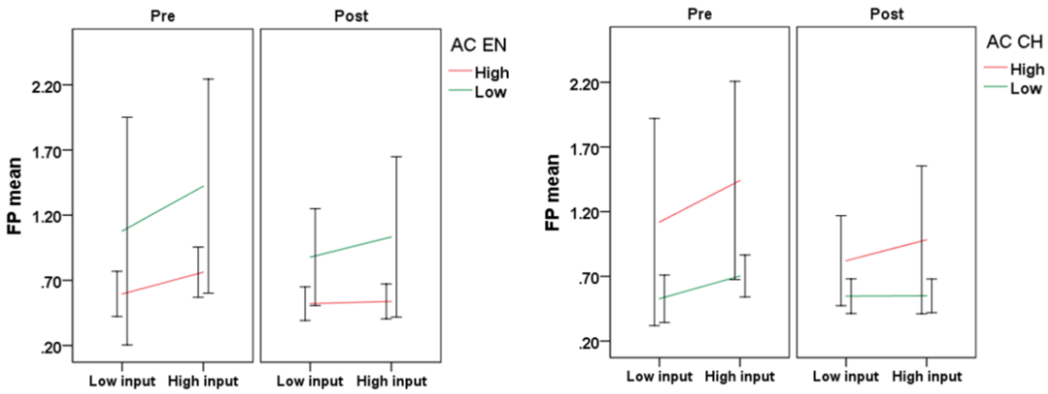
The effects of cognitive fluency, training and input rate on FP mean are illustrated below in the visualised charts, and the changes in FP mean in the interpreted output are discussed.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.17 Influence of LA, LR, training and input rate on FP mean

As shown in Chart 6.17, the FP mean of the high LA EN group was generally smaller than that of the low LA EN group under both low and high input rate conditions in both the pre-test and post-test. The FP mean under high input rate conditions was larger than that under low input rate conditions, except for the high LA EN group in the post-test, where the FP mean under conditions of low and high input rates was similar. The FP mean under conditions of high input rate increased for the low LA EN group but decreased for the high LA EN group, which implies that the high LA EN group might have had better control of filled pauses under high cognitive load. The pattern for the LA CH groups was very similar to that of the LA EN groups. In terms of LR, there were more filled pauses in the SI output for the high LR efficiency group compared to the low LR efficiency group. The FP mean under conditions of high input rate was slightly larger than that under low input rate conditions for both LR groups in both the pre-test and the post-test.

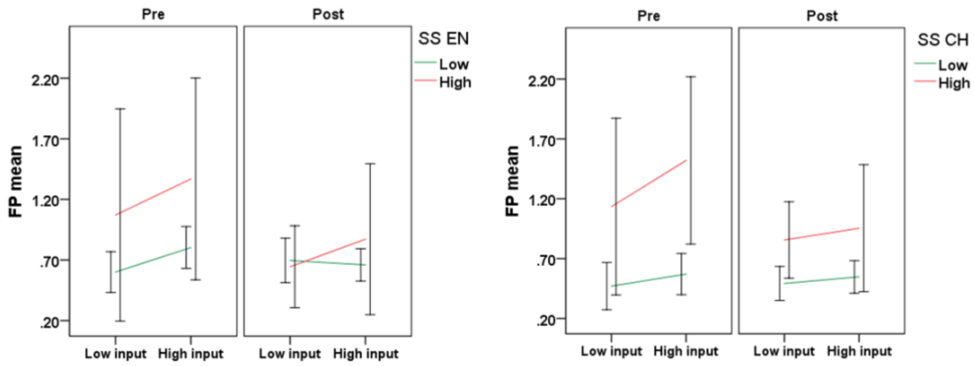


Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.18 Influence of AC, training and input rate on FP mean

For AC efficiency, the FP mean of the high AC EN group was smaller than that of the low AC EN group under conditions of both low and high input rates in both the pre-test and the post-test. In terms of AC CH efficiency, however, the trend was reversed, and the FP mean of the high AC CH group was larger than that of the low AC CH group under conditions of both low and high input rates in both the pre-test and the post-test. L1 and L2 efficiency of AC seem to have a different impact on FP mean, though the differences in FP mean between low and high AC groups did not reach significance.



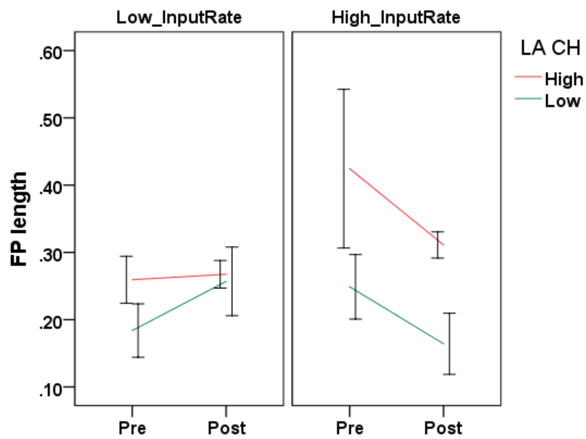
Notes: — High represents a high capacity of working memory;
— Low represents a low capacity of working memory.

Chart 6.19 Influence of SS, training and input rate on FP mean

As illustrated in Chart 6.19, the FP mean of the high SS EN group was generally larger than that of the low SS EN group, except under low input rate conditions in the post-test. There were more filled pauses in the high SS CH group compared with the low SS CH group under conditions of low and high input rates in both the pre-test and the post-test. Generally, the high SS EN and SS CH groups tended to have more filled pauses in their SI output, though the differences between the low and high SS groups were not significant.

6.1.2.4 Effects of cognitive fluency, training and input rate on FP length

Repeated measures ANOVA analyses were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on FP length (mean duration of filled pauses). The results of the analyses showed that the main effect of LA CH on FP length was significant [$F(1,26) = 4.446$, $p = 0.045 < 0.05$], but the effects of all other cognitive fluency measures on FP length were insignificant. The pairwise comparison showed that the FP length of the high LA CH efficiency group was longer than that of the low LA CH group by 0.102. Neither SI training nor input rate had a significant impact on FP length. There were no interaction effects between cognitive fluency measures, training and input rate. As illustrated in Chart 6.20, the FP length of the high LA CH efficiency group was larger than that of the low LA CH group under conditions of both low and high input rates in both the pre-test and the post-test. The chart also shows that in the post-test, the difference in FP length between the low and high LA CH groups was very small under low input rate conditions, but the difference was quite apparent under high input rate conditions.



Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.20 Influence of LA CH, training and input rate on FP length

6.1.3 Effects of cognitive fluency, training and input rate on REPAIR fluency in SI

Results of repeated measures ANOVA analyses found that indicators of repair fluency in the SI output were significantly affected by one or more measures of cognitive fluency, SI training and/or input rate of source speeches. Their effects on REPAIR mean and REPAIR length are discussed in detail in this section.

6.1.3.1 Effects of cognitive fluency, training and input rate on REPAIR mean

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on REPAIR mean (mean number of the sum of repair, fillers and repetitions per minute).

In terms of cognitive fluency, the effect of SS CH on REPAIR mean approached significance [$F(1,26)=4.208$, $p=0.050$] and the effects of other cognitive fluency measures on REPAIR mean did not reach significance. The pairwise comparison showed that the REPAIR mean of the high SS CH group was larger than that of the low SS CH group by 0.701, which implied that participants with higher target language working memory capacity had more REPAIRs in their SI output.

Interaction effects were found between training and input rate. Results of analyses for the interaction effects and simple effects of training and input rate are summarized in the Table 6.7.

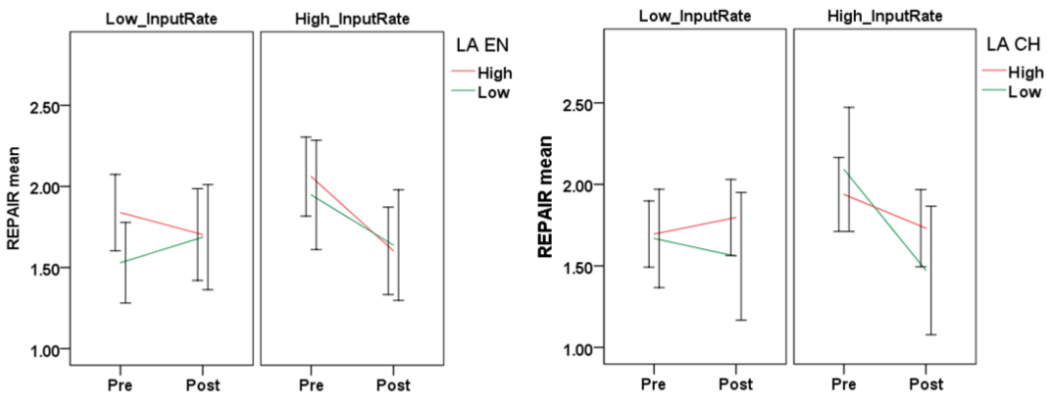
Table 6.7 Summary of results for the interaction and simple effects of training and input rate on REPAIR mean

	Between-subject Variables	**Interaction Effects**	**Simple Effects**	
		Training*Input rate	**Training Effects _High input rate**	**Input Rate Effects _Pre-test**
REPAIR mean	LA EN	* [F(1,26)=5.041, p=0.033]	*[F(1,26)=5.578, p=0.026]	* [F(1,26)=6.369, p=0.018]
	LA CH	* [F(1,26)=5.337, p=0.029]	*[F(1,26)=6.688, p=0.016]	* [F(1,26)=6.727, p=0.015]
	LR			
	AC EN	* [F(1,26)=6.376, p=0.018]	*[F(1,26)=6.414, p=0.018]	**[F(1,26)=9.082, p=0.006]
	AC CH	* [F(1,26)=5.281, p=0.030]	*[F(1,26)=5.904, p=0.022]	* [F(1,26)=6.744, p=0.015]
	SS EN	* [F(1,26)=5.75, p=0.024]	*[F(1,26)=7.142, p=0.013]	**[F(1,26)=8.040, p=0.009]
	SS CH	*[F(1,26)=5.248, p=0.030]	*[F(1,26)=5.875, p=0.023]	* [F(1,26)=6.941, p=0.014]

Note: * p＜0.05; ** p＜0.01

The two-way interaction between training and input rate were significant except when with LR efficiency as the between-subjects variable. There were no other interaction effects among cognitive fluency, SI training and input rate. Further analyses of simple effects showed that: The training effects were only significant for the high input rate conditions where there were more REPAIRs in the pre-test than in the post-test, but the training effects were not significant under low input rate conditions. The effects of input rate were only significant in the pre-test, but not in the post-test. There were more REPAIRs under conditions of high input rate compared with that under low input rate in the pre-test.

The effects of cognitive fluency, training and input rate on REPAIR mean are illustrated below based on the visualised charts, and the changes in REPAIR mean in the SI output are discussed.

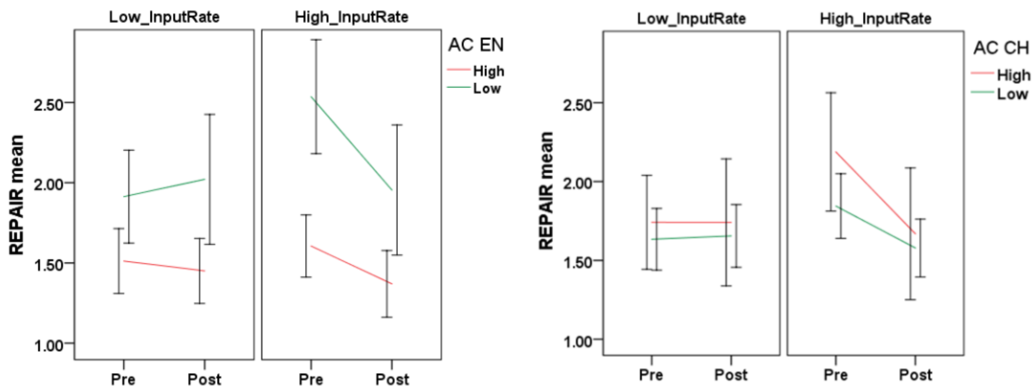


Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.21 Influence of LA, training and input rate on REPAIR mean

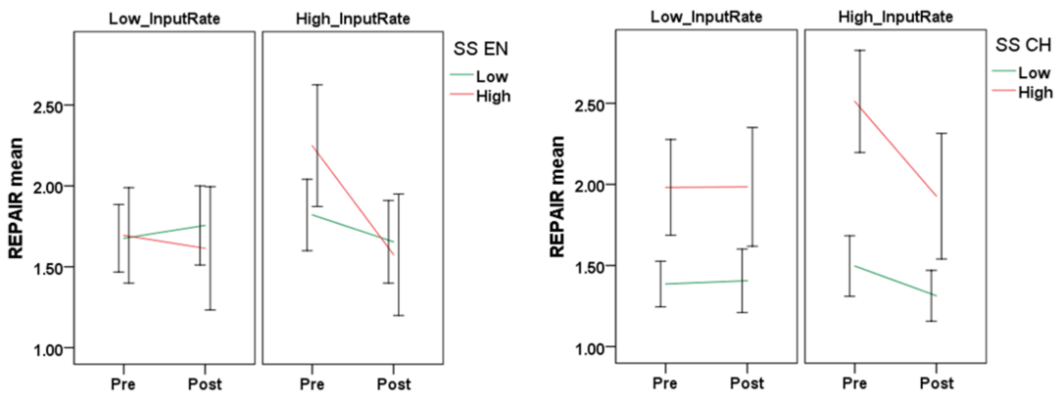
As illustrated in Chart 6.21, the REPAIR mean in the post-test was significantly smaller than that in the pre-test under high input rate conditions for both the LA EN and LA CH groups. In terms of the impact of input rate, REPAIR mean under the high input rate condition was more than that under the low input rate condition in the pre-test for both the LA EN and LA CH groups. But the REPAIR mean under low and high input rate conditions was similar for both LA EN and LA CH groups in the post-test.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.22 Influence of AC, training and input rate on REPAIR mean

The REPAIR mean in the post-test was smaller than that in the pre-test under high input rate conditions for both the AC EN and AC CH groups. Under low input rate conditions, there was only a slight change in the REPAIR mean for both the AC EN and AC CH groups after training. In terms of the impact of input rate, the REPAIR mean under high input rate conditions was larger than that under low input rate conditions in the pre-test for the low LA EN and high LA CH groups, while the mean number was similar for the high LA EN and low LA CH groups in the pre-test. Additionally, the REPAIR mean under conditions of low and high input rates was similar for each of the LA EN and LA CH groups in the post-test.



Notes: —— High represents a high capacity of working memory;

—— Low represents a low capacity of working memory.

Chart 6.23 Influence of SS, training and input rate on REPAIR mean

As shown in Chart 6.23, the REPAIR mean of the high SS CH group was much larger than that of the low SS CH group under conditions of both low and high input rates in both the pre-test and the post-test. The REPAIR mean of the high SS EN group was generally smaller after training under conditions of both low and high input rates and the change was more evident under high input rate conditions. For the low SS EN group, there were slightly more REPAIRs under low input rate conditions and fewer REPAIRs under high input rate conditions, but the change was minimal and not significant.

The REPAIR mean in the post-test was smaller than that in the pre-test under conditions of high input rate for both SS EN and SS CH groups. Under conditions of low input rate, the change in REPAIR mean was slight for the SS EN and SS CH groups. In terms of the impact of input rate, the REPAIR mean under high input rate conditions was greater than that under conditions of low input rate in the pre-test for the high SS EN and high SS CH groups, whereas the numbers remained similar for the low SS EN and low SS CH groups. However, the REPAIR mean under conditions of low and high input rates was similar for each of the SS EN and SS CH groups in the post-test.

6.1.3.2 Effects of cognitive fluency, training and input rate on REPAIR length

Repeated measures ANOVA were performed to analyse the impact of cognitive fluency measures (LA EN, LA CH, LR, AC EN, AC CH, SS EN, SS CH), SI training and input rate on REPAIR length (mean duration of repairs, repetitions and false starts). Results are summarized in Table 6.8.

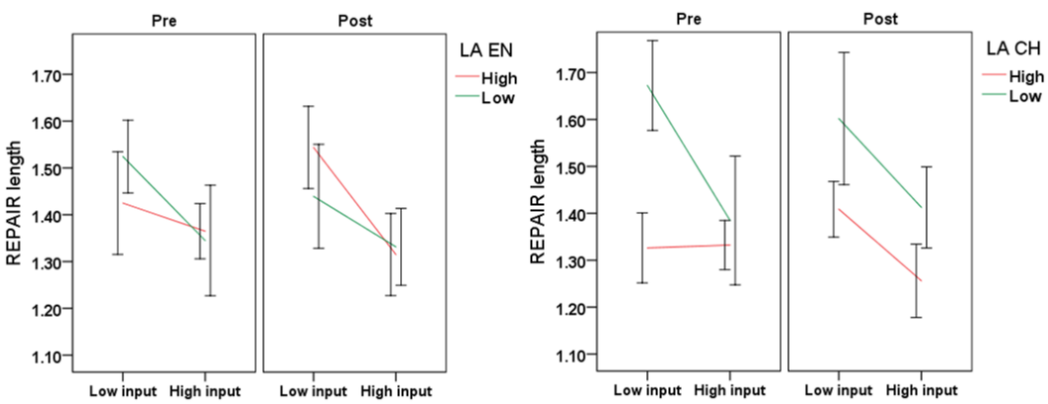
Table 6.8 Summary of results for the effects of cognitive fluency and input rate on REPAIR length

	Between-subject Variables	Main Effects	
		Cognitive Fluency	**Input Rate**
REPAIR length	LA EN		◐ [F(1,26)=3.958, p=.057]
	LA CH	* [F(1,26)=6.755, p=.015]	* [F(1,26)=4.773, p=.038]
	LR		
	AC EN		
	AC CH	* [F(1,26)=4.457, p=.045]	◐ [F(1,26)=4.083, p=.054]
	SS EN		* [F(1,26)=4.972, p=.035]
	SS CH		◐ [F(1,26)=3.962, p=.057]

Note: * p＜0.05; ◐ approaching significance.

The effects of LA CH and AC CH on REPAIR length were significant and the effects of other cognitive fluency measures on REPAIR length did not reach significance. Pairwise comparisons showed that: The REPAIR length of the high LA CH efficiency group was shorter than that of the low LA CH efficiency group by 0.187. The REPAIR length of the high AC CH group was longer than that of the low AC CH group by 0.156. It indicates that there was longer duration of REPAIRs for participants with lower efficiency of LA CH or higher efficiency of AC CH. The effects of input rate on REPAIR length were significant with LA CH and SS EN as the between-subjects variables and approached significance with LA EN, AC CH or SS CH as the between-subjects variables. It indicates that there was a significant or nearly significant change in the REPAIR length under conditions of low and high input rates for groups of LA EN, LA CH, AC CH, SS EN and SS CH. Pairwise comparisons showed that the REPAIR length under low input rate conditions was longer than that under high input rate conditions. The input rate effects were not significant for groups of LR or AC EN. The effects of training on REPAIR length did not reach significance, which indicates that there was no significant change in the REPAIR length between the pre-test and the post-test. Besides, there were no significant interaction effects among cognitive fluency, SI training and input rate.

The effects of cognitive fluency, training and input rate on REPAIR length are illustrated below based on the visualised charts, and the changes in REPAIR length in the interpreted output are discussed.

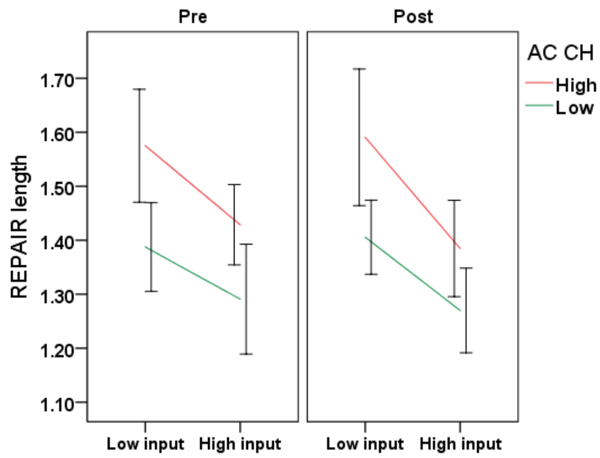


Notes: —— High represents high efficiency of the cognitive process;

—— Low represents low efficiency of the cognitive process.

Chart 6.24 Influence of LA, training and input rate on REPAIR length

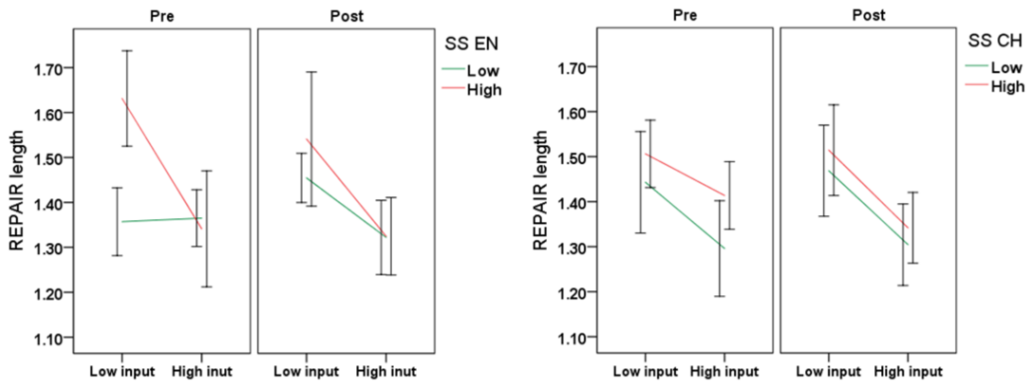
As shown in Chart 6.24, the REPAIR length of the high LA CH group was shorter than that of the low LA CH group under conditions of both low and high input rates in both the pre-test and the post-test. For both LA EN and LA CH groups, the REPAIR length under the high input rate condition was generally shorter than that under the low input rate condition in both the pre-test and the post-test, except for the high LA CH group in the pre-test, for which the REPAIR length was similar under conditions of low and high input rates.



Notes: —— High represents high efficiency of the cognitive process;
—— Low represents low efficiency of the cognitive process.

Chart 6.25 Influence of AC CH, training and input rate on REPAIR length

As illustrated in Chart 6.25, the REPAIR length of the high AC CH group was longer than that of the low AC CH group under conditions of both low and high input rates in both the pre-test and the post-test and the differences were significant. For both the low and high AC CH groups, the REPAIR length was shorter under high input rate conditions than under conditions of low input rate in both the pre-test and the post-test.



Notes: —— High represents a high capacity of working memory;

—— Low represents a low capacity of working memory.

Chart 6.26 Influence of SS, training and input rate on REPAIR length

As shown in Chart 6.26, the REPAIR length of the high SS EN group was longer than that of the low SS EN group under low input rate conditions in both the pre-test and the post-test, but the REPAIRs were of similar duration under high input rate conditions for both low and high SS EN groups in the pre-test and the post-test. The REPAIR length of the high SS CH group was longer than that of the low SS CH group under conditions of both low and high input rates in both the pre-test and the post-test, but the differences between the low and high SS groups did not reach significance.

For both SS EN and SS CH groups, the REPAIR length under conditions of high input rate was generally shorter than that under low input rate conditions in both the pre-test and the post-test, except for the low SS EN group in the pre-test, for which the REPAIR length was similar under low and high input rates.

6.2 Summary and discussion of ANOVA results

This section summarises and discusses results of the three-factor repeated measures ANOVA analyses of the effects of cognitive fluency, SI training and input rate on utterance fluency indicators in the SI performance of trainee interpreters. The ANOVA analyses were performed with SI training and input rate as within-subjects variables and each of the cognitive fluency measures as the between-subjects variable. Two-way interaction effects are discussed, followed by the main effects of cognitive fluency, SI training and input rate on indicators of utterance fluency in SI output.

6.2.1 Interaction effects between cognitive fluency, training and input rate

Analyses of the interaction of cognitive fluency, training and input rate are important for exploring the interactive role of these three factors in utterance fluency measures in trainee interpreters' SI output. Two-way interaction effects between input rate and cognitive fluency and between training and input rate in the ANOVA analyses are summarised in Table 6.9.

Table 6.9 Summary of significant interaction effects from ANOVA analyses

Fluency Indicators	**Interaction Effects**						
	LA EN	LA CH	LR	AC EN	AC CH	SS EN	SS CH
SR						●	
AR	●	●	●				
MLR						●	
SP mean					●		
REPAIR	▲	▲		▲	▲	▲	▲

Notes: ● Significant two-way interaction (input rate*cognitive fluency), $p < 0.05$
▲ Significant two-way interaction (training*input rate), $p < 0.05$

Table 6.10 Summary of results for the simple effects in two-way interactions

<table>
<tr><th>Interaction Effects</th><th>Fluency Indicators</th><th colspan="3">Simple Effects of Input Rate</th><th colspan="3">Simple Effects of Training/Cognitive Fluency</th></tr>
<tr><td rowspan="2">Training *Input▲</td><td rowspan="2">REPAIR mean</td><td rowspan="2">Input</td><td>Pre-test</td><td>*↑</td><td rowspan="2">Training</td><td>high input</td><td>*↓</td></tr>
<tr><td>Post-test</td><td></td><td>low input</td><td></td></tr>
<tr><td rowspan="12">Input* Cognitive fluency
●</td><td rowspan="2">SR</td><td rowspan="2">Input</td><td>high SS EN</td><td>*↑</td><td rowspan="2">SS EN</td><td>high input</td><td>*↑</td></tr>
<tr><td>low SS EN</td><td>◐↑</td><td>low input</td><td>x</td></tr>
<tr><td rowspan="2">AR</td><td rowspan="2">Input</td><td>high LA CH</td><td>*↑</td><td rowspan="2">LA CH</td><td>high input</td><td>x</td></tr>
<tr><td>low LA CH</td><td>*↑</td><td>low input</td><td>x</td></tr>
<tr><td rowspan="2">AR</td><td rowspan="2">Input</td><td>high LA EN</td><td>*↑</td><td rowspan="2">LA EN</td><td>high input</td><td>x</td></tr>
<tr><td>low LA EN</td><td>*↑</td><td>low input</td><td>x</td></tr>
<tr><td rowspan="2">AR</td><td rowspan="2">Input</td><td>high LR</td><td>*↑</td><td rowspan="2">LR</td><td>high input</td><td>x</td></tr>
<tr><td>low LR</td><td>*↑</td><td>low input</td><td>x</td></tr>
<tr><td rowspan="2">MLR</td><td rowspan="2">Input</td><td>high SS EN</td><td>*↑</td><td rowspan="2">SS EN</td><td>high input</td><td>x</td></tr>
<tr><td>low SS EN</td><td>*↑</td><td>low input</td><td>x</td></tr>
<tr><td rowspan="2">SP mean</td><td rowspan="2">Input</td><td>high AC CH</td><td>*↓</td><td rowspan="2">AC CH</td><td>high input</td><td>x</td></tr>
<tr><td>low AC CH</td><td>x</td><td>low input</td><td>x</td></tr>
</table>

Notes: * significant, $p < 0.05$; ◐ approaching significance; *x* not significant;
↑ larger value under high input rate/in post-test/for high cognitive fluency group;
↓ smaller value under high input rate/in post-test/for high cognitive fluency group.

Significant two-way interaction effects between training and input were found only for REPAIR mean. Significant two-way interaction effects between cognitive fluency and input were found for fluency indicators including SR, AR, MLR and SP mean. The results of further analyses of simple effects in the interaction effects are summarised in Table 6.10 and discussed below.

For the interaction of training and input rate, the training effects were only significant under high input rate conditions where there were fewer REPAIRs in the post-test than in the pre-test. The effects of input rate were only significant in the pre-test when there were more REPAIRs under high input rate than under low input rate. This indicates that the trainee interpreters were significantly affected by high input rate in the pre-test, but not in the post-test, in which they managed to cope with high cognitive load and had better control of REPAIRs after training with the development of interpreting expertise.

For the interaction of input rate and cognitive fluency, significant effects were found between input rate and SS EN for SR, between input rate and LA CH, LA EN, LR separately for AR, between input rate and SS EN for MLR, and between input rate and AC CH for SP mean. For main effects of cognitive fluency, only the effects of SS EN were significant under high input rate conditions for SR, with the SR of the high SS EN group faster under conditions of high input rate. Other effects of cognitive fluency did not reach significance, as listed in Table 6.11.

The main effects of input rate were prevalent for nearly all listed cognitive fluency groups, except the low AC CH group for SP mean. The SR and AR under high input rate conditions were significantly faster than under low input rate conditions for correspondent low and high cognitive fluency groups (SS EN, LA CH, LA EN and LR respectively). The MLR under high input rate conditions was significantly longer than that under low input rate conditions for both low and high SS EN groups. There were significantly fewer silent pauses per minute under high input rate conditions compared with conditions of low input rate for the high AC CH group.

6.2.2 Main effects of cognitive fluency

The effects of LA CH, LR, AC CH and SS CH on one or more SI fluency indicators were significant, which indicates that there was a significant difference in correspondent fluency indicators between these low and high cognitive fluency groups, as shown in Table 6.11. This provides further evidence of the important role of cognitive fluency in utterance fluency of trainee interpreters' SI output.

Table 6.11 Summary of results for significant main effects of cognitive fluency

Fluency Indicators		Main Effects of Cognitive Fluency			
		LA CH	LR	AC CH	SS CH
Speed Fluency	SR		●		
	MLR		●		
	PTR		●		
Breakdown Fluency	SP length		◐		
	FP length	●			
Repair Fluency	REPAIR mean				◐
	REPAIR length	●		●	

Notes: ● represents significant main effects, $p < 0.05$;

◐ represents main effects approaching significance, $p \approx 0.05$

For cognitive fluency measured in both in L1 (CH) and L2 (EN) including the LA, AC and SS measures, only the L1 (the target language in SI) measures had significant or near-significant impacts on one or more indicators of SI utterance fluency, while cognitive fluency measures in L2 (source language in SI) had no significant impacts. This indicates that the efficiency of target language processing had a more significant impact on SI fluency than that of source language processing. This is consistent with the results of analyses discussed in the previous chapter that L1 working memory span has a significant impact on the development of SI utterance fluency while L2 working memory span does not. It is a further proof of the domain-specific view of working memory (Miyake, 2001; Shah & Miyake, 1996).

The main effects of LA CH on FP length and REPAIR length were significant, with different impacts. The FP length of the high LA CH efficiency group was longer, while that group's REPAIR length was shorter than that of the low LA CH group, under conditions of both low and high input rate in both the pre-test and the post-test. Filled pauses and REPAIRs are both regarded as disfluencies and mirror certain cognitive difficulties in SI. The filled pause, as a kind of hesitation pause, reflects planning difficulties caused by either conceptualisation or information formulation in SI. According to the gravitational model (Gile, 2009), high production availability would accelerate the production process. Though the overall pace of the SI output is determined by the source speech, higher production efficiency would allow the interpreters more leeway for planning and save some processing capacity for comprehension and coordination, which is a possible reason for the longer FP length for the high LA CH group. Due to the high cognitive load in SI, it costs much additional cognitive effort for interpreters to make REPAIRs (repairs, repetitions or correcting a false start). Higher efficiency in the target language may have allowed the trainee interpreters to make these REPAIRs more efficiently and quickly, leading to a shorter duration of REPAIRs.

The main effects of LR on SR, MLR and PTR were significant and its effect on SP length approached significance. The SI output of the high LR efficiency group had faster SR, longer MLR, higher PTR and shorter mean duration of silent pauses, compared to those of the low LR efficiency group. The LR measures elicited from the oral word translation task in this research also reflect the efficiency of language conversion and verbal articulation. This indicates that trainee interpreters with high LR efficiency managed to produce more and longer utterances in their SI output with faster speed, which at the same time led to shorter silent pausing time. This finding provides evidence for the significant impact of LR on the utterance fluency of trainee interpreters' SI output, though its contribution to the development of SI utterance fluency is limited based on the results of analyses in the previous chapter. It also consolidates Christoffels et al.'s (2003) finding that retrieval of translation equivalence is an important contributor to SI performance.

The AC CH had a significant influence on REPAIR length and the SI output of the high AC CH efficiency group exhibited a longer duration of REPAIRs under conditions of both low and high input rates and in both the pre-test and the post-test. The high AC CH group had a smaller shift cost, which represents a smaller burden on the interpreters when shifting the focus of language-directed attention. When REPAIRs are made, the comprehension of the incoming message and monitoring of the interpreter's own production all require adequate attention from the interpreters. A possible explanation is that interpreters with high AC CH might have been more efficiently directed by the language processing demand in SI and devoted more cognitive effort to processes like the comprehension of incoming information since it costs them a smaller burden compared to the low AC CH group. Due to the complexity of these cognitive processes, however, further research is required to verify the role of language-directed attention control in SI.

The SS CH had a significant influence on REPAIR mean. There were more REPAIRs per minute in the SI output of the high SS CH group compared with that of the low SS CH group for conditions of both low and high input rates in both the pre-test and the post-test. While making REPAIRs, interpreters continue to store incoming messages while processing the current segment of information. Larger working memory capacity in the target language allows more information to be stored before the overload of processing capacities occurs. It also makes REPAIRs possible since interpreters would only make REPAIRs when they have extra processing abilities to process the still-incoming information. This further provides evidence for the role of target language working memory in SI performance.

Cognitive fluency overall plays an independent role in utterance fluency of the SI output of trainee interpreters. The simple effects of cognitive fluency in the two-way interaction between input rate and cognitive fluency did not reach significance in most cases, with only one exception of the significant impact of SS EN on SR under the high input rate condition. Moreover, the above discussion indicates that a higher cognitive fluency efficiency is not necessarily conducive to more fluent SI output. Trainee interpreters with high lexical retrieval efficiency may produce a more fluent interpreted output, indicated by a faster SR, longer MLR, higher PTR and shorter silent pausing time. The efficiency of LA, AC or SS does not necessarily contribute to SI utterance fluency, but the impact might be positive for overall SI performance, which deserves further investigation.

6.2.3 Main effects of SI training

The main effects of SI training were significant for SR, AR and SP length for each of the individual cognitive fluency measures as the between-subject variables, but not for other fluency indicators including SP mean, FP mean, FP length, REPAIR mean and REPAIR length, as summarised in Table 6.12.

Table 6.12 Summary of results for significant main effects of training

Fluency Indicators	Main Effects of Training						
	LA EN	LA	LR	AC	AC	SS EN	SS CH
SR	**	*	**	*	*	**	**
AR	*	*	*	*	**	*	*
SP length	**	**	**	**	**	**	**

Note: * $p < 0.05$; ** $p < 0.01$

The SR and AR of the interpreted output were higher after SI training and the mean duration of silent pauses was longer after SI training. Trainee interpreters were able to execute target language production at faster speeds after a certain period of SI training. The development of SI expertise and the better use of interpreting strategies after SI training might explain this. The more efficient target language production would allow the trainee interpreters more time for information planning as indicated by longer silent pausing time, though the overall SI process is paced by the source speech.

Moreover, analyses of the simple effects for the interaction of training and input rate showed that the training effects were significant under conditions of high input rate where there were on average fewer REPAIRs in the post-test, but the training effects were not significant for REPAIR mean under low input rate conditions. This indicates that trainee interpreters could cope with conditions of high cognitive load better after a period of SI training.

The SI output after training tended to be more fluent than that before training, as indicated by faster SR and AR, longer silent pausing time and fewer REPAIRs under conditions of high cognitive load. It must be noted that the training period in the current research was comparatively short at only thirteen weeks, though the training was intensive. The training effects were probably limited for this short period and might be more apparent with longer SI training time. Future longitudinal research with a longer period of SI training on the effects of interpreting training on SI utterance fluency is needed.

6.2.4 Main effects of input rate

The main effects of input rate were strongly significant for speed fluency indicators including SR, AR and MLR, overall significant for the breakdown fluency indicators of SP mean and FP mean, and partly significant for the repair fluency indicator of REPAIR length in the SI output, as shown in Table 6.13.

Table 6.13 Summary of results for significant main effects of input rate

Fluency Indicators	**Main Effects of Input Rate**						
	LA EN	LA CH	LR	AC EN	AC CH	SS EN	SS CH
SR	**	**	**	**	**	**	**
AR	**	**	**	**	**	**	**
MLR	**	**	**	**	**	**	**
SP mean	*	*		*	*	*	*
FP mean	*	*	*	*	*	*	*
REPAIR length	◐	*			◐	*	◐

Note: * $p<0.05$; ** $p<0.01$; ◐ $p\approx0.05$

The SI output under high input rate conditions exhibited faster SR and AR and longer MLR, and had fewer silent pauses but more filled pauses, compared with that under conditions of low input rate. In terms of repair fluency indicators, the input rate only affected the mean duration of REPAIRs significantly for some of the cognitive fluency groups. The input rate only had a significant impact on LA CH and SS EN groups and the input rate effects approached significance for the LA EN, AC CH and SS CH groups.

The mean duration of REPAIRs was shorter under high input rate conditions compared with that under conditions of low input rate in these cases. Furthermore, the effects of input rate were only significant in the pre-test for REPAIR mean, but not in the post-test.

Previous studies have found strong correlations between objective utterance fluency measures and perceived fluency. Speed fluency indicators were found to be positively related to rater-generated ratings while the number of silent pauses and repair fluency indicators were negatively related to overall fluency ratings (Han, 2015a; Yu & van Heuven, 2017). With higher SR and AR, longer MLR, fewer silent pauses and shorter duration of REPAIRs, the SI output of trainee interpreters under conditions of high input rate was generally more fluent than that under low input rate conditions, though more filled pauses were found when input rate was high. This finding is inconsistent with those of previous studies (e.g. Gerver, 1969/2002; Pio, 2003) which found that high input rate had a negative impact on SI performance, including the dimension of fluency. However, it provides additional evidence for the view that interpreters' SI output might be more fluent under high input rate conditions, which is consistent with Han and Riazi's (2017) findings with professional interpreters as participants. It indicates that trainee interpreters have developed a certain level of SI expertise after a period of SI training to cope with high cognitive load, which has implications for interpreting pedagogy and assessment.

One of the challenges in interpreter training is choosing appropriate materials for training and assessment for each stage of the training. Incremental realism is one of the key principles of interpreter training, which means "moving from targeted 'quasiinterpreting' exercises through artificially structured, then more authentic, realistic speeches progressing in difficulty on multiple parameters (register, subject matter, speed, form of presentation, etc.), as well as in expectations of the product" (Setton & Dawrant, 2016, p. 30). Delivery speed is an important difficulty index of the source speech. In practice, trainers often tailor the speech material by slowing it down. The finding that trainee interpreters' SI output (from B to A language) tends to be more fluent under high input rate conditions after a period of training sheds light on interpreting pedagogy. It is worth considering that trainers should adjust and diversify their interpreting training material based on the level of expertise development of trainee interpreters even at the early stage of SI training. It should be noted that this does not contradict the general guiding principle of incremental realism. The addition of some practice with high input rate material may have beneficial effects, for instance, the development of cognitive control abilities. It must be admitted, however, that the enhanced fluency might be at the cost of information accuracy. The value of this potential method should not be denied, though further empirical research is required to verify the finding.

Chapter 7 The relationship between SI utterance fluency and perceived fluency

This chapter explores the relationship between utterance fluency and perceived fluency of trainee interpreters' SI performance. A series of analyses were conducted to explore to what extent could perceived fluency, as indicated by fluency ratings, be predicted from objectively calculated indicators of utterance fluency in trainee interpreters' SI output. With the role of cognitive fluency in the development of utterance fluency and the impact of cognitive fluency on utterance fluency confirmed in previous chapters, the investigation into the relationship between utterance and perceived fluency could lead to a better understanding of the multidimensionality of fluency in SI.

7.1 Data analysis

7.1.1 Rating reliability

Reliability refers to the consistency of measurement, i.e. performances of similar quality receiving similar scores (Lee, 2008). Inter-rater consistency (the consistency between different raters) and intra-rater consistency (the consistency within the same rater across rating occasions) are used to ensure the reliability of assessment (Bachman, 1990). Inter-rater reliability is the agreement between raters on the same assessment, i.e. raters giving the same or similar scores to the same performance (Sawyer, 2004). Since the current research did not undertake a double rating or re-testing procedure, intra-rater reliability was not examined.

In this research, inter-rater reliability was measured using Kendall's coefficient of concordance (W), a simple and efficient measure for assessing agreement between observers (Gearhart, Booth, Sedivec, & Schauer, 2013; Kendall & Babington Smith, 1939). Kendall's W ranges from 0 (complete disagreement) to 1 (perfect agreement). The agreement is regarded as strong if Kendall's W is within the 0.71–0.90 rank and very strong if it is above 0.9 (LeBreton & Senter, 2008).

Inter-rater reliability was examined separately for each of the four conditions (Pre_S, Pre_F, Post_S and Post_F). The results showed that the three raters' ratings for fluency and overall SI performance were concordant with one another and the agreement was strong. The results of analyses showed that the global concordance (W) values among raters' fluency ratings ranged from 0.73 to 0.79 (*Kendall's* $W = 0.730$, $p < 0.001$ for Pre_S, *Kendall's* $W = 0.758$, $p < 0.001$ for Pre_F, *Kendall's* $W = 0.790$, $p < 0.001$ for Post_S and *Kendall's* $W = 0.737$, $p < 0.001$ for Post_F). The global concordance (W) values among raters' overall SI performance ratings ranged from 0.701 to 0.734 (*Kendall's* $W = 0.714$, $p = 0.001$ for Pre_S, *Kendall's* $W = 0.739$, $p < 0.001$ for Pre_F, *Kendall's* $W = 0.734$, $p < 0.001$ for Post_S and *Kendall's* $W = 0.701$, $p = 0.001$ for Post_F). All p values were significant at the 0.05 level.

7.1.2 Predicting perceived fluency from utterance fluency measures

To explore the predicting role of objectively measured utterance fluency indicators in perceived fluency as indicated by human ratings of trainee interpreters' SI performance, multiple linear regression analyses were performed for each of the four conditions, i.e. the low input rate condition in the pre-training task (Pre_S), the high input rate condition in the pre-training task (Pre_F), the low input rate condition in the post-training task (Post_S) and the high input rate condition in the post-training task (Post_F) separately. The average score of the three raters was adopted as the measure for perceived fluency since the agreement among different raters was strong, as indicated by Kendall's coefficient of concordance in the above analyses. Results of relevant analyses are presented and discussed in this section.

Correlations between the objectively measured utterance fluency indicators and fluency ratings were first analysed in order to explore which of the quantitative variables might be successful predictors of fluency ratings. Pearson correlation analyses indicated that SR, PTR, MLR and SP length correlated significantly ($p < 0.05$) with fluency ratings under all four conditions (Pre_S, Pre_F, Post_S and Post_F) and AR had a significant correlation with fluency ratings under the high input rate condition in the pre-training task ($r = 0.410$, $p < 0.05$). It was noted that SR had the strongest correlation with fluency ratings. Utterance fluency indicators which had a significant correlation with fluency ratings are listed in Table 7.1. The correlations between fluency ratings and SP mean, FP mean, FP length, REPAIR mean and REPAIR length did not reach significance.

Table 7.1 Correlations between utterance fluency indicators and fluency ratings

Fluency Rating	**SR**	**AR**	**PTR**	**MLR**	**SP length**
Pre_S	.717**	.307	.443*	.564**	-.468*
Pre_F	.737**	.410*	.450*	.576**	-.455*
Post_S	.699**	.329	.548**	.593**	-.545**
Post_F	.749**	.154	.559**	.603**	-.562**

Notes: ** significant at the 0.01 level (2-tailed);
* significant at the 0.05 level (2-tailed).

Multiple linear regression analyses were then conducted to further explore if a combination of independent variables would allow better predictions, taking objective measures of utterance fluency of trainee interpreters' SI output as the independent variables and fluency ratings as the dependent variables. Independent and dependent variables of the regression analyses are listed in Table 7.2. The correlations of the ten indicators of utterance fluency were analysed and the results showed that there were comparatively strong correlations between several of the independent variables (SR and PTR, SR and SP length, PTR and SP length). Stepwise multiple linear regressions were conducted to select predictors and to estimate the unique contribution of each independent variable in predicting fluency ratings. Stepwise multiple regression analysis is used for prediction, which can help determine a useful subset of variables for predicting outcomes (Keith, 2015). In stepwise regression, predictor variables are entered one at a time in sequential order. "The order of entry of the variables is controlled by the statistics program; the variable that will lead to the largest increase in ΔR_2 is entered at each step. If an earlier variable becomes statistically not significant with the addition of later variables, it can be dropped from the equation" (ibid., p. 101).

Table 7.2 Independent and dependent variables of stepwise multiple linear regression analyses

Independent Variables (Utterance Fluency)	Speed fluency	SR	speech rate
		AR	articulation rate
		PTR	phonation time ratio
		MLR	mean length of run
	Breakdown fluency	SP mean	mean number of silent pauses per minute
		SP length	mean duration of silent pauses
		FP mean	mean number of filled pauses per minute
		FP length	mean duration of filled pauses
	Repair fluency	REPAIR mean	mean number of REPAIRs (repairs, repetitions and false starts) per minute
		RP length	mean duration of REPAIRs (repairs, repetitions and false starts)
Dependent Variables	Fluency ratings of trainee interpreters’ SI performance in the Pre_S, Pre_F, Post_S and Post_F tasks separately		

7.1.2.1 Predicting role of utterance fluency indicators in Pre_S

Stepwise multiple linear regression analysis was conducted using the fluency rating in the Pre_S task (the pre-training task under conditions of low input rate) as the dependent variable and utterance fluency indicators as listed in Table 7.2 in the Pre_S SI performance as independent variables. The stepwise multiple linear regression model was statistically significant [$F(2,27) = 22.702$, $p < 0.001$, $R^2 = 0.645$, *adjusted* $R^2 = 0.616$]. Results of the analysis are presented in Table 7.3.

Table 7.3 Coefficients of the regression model for Pre_S fluency rating

Model	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
(Constant)	61.268	3.628		16.887	.000		
Pre_S_SR	8.054	1.273	.760	6.329	.000	.986	1.014
Pre_S_REPAIRmean	-1.627	.535	-.365	-3.040	.005	.986	1.014

The independent variables contributing to predicting the dependent variable were SR ($\beta = 0.760$, $t = 6.329$, $p < 0.001$) and REPAIR mean ($\beta = -0.365$, $t = -3.040$, $p = 0.005 < 0.01$). The remaining predictors were removed from the model. The results of analyses showed that the variable that explained the greatest amount of variance of the dependent variable was SR, accounting for 51.4% of the variance of the dependent variable alone, and its correlation with the dependent variable was significantly positive with a coefficient of 8.054. The second variable added into the model was REPAIR mean, increasing the explained variance to 64.5%, and its correlation with the dependent variable was significantly negative with a coefficient of -1.627. The absolute value of the standardised regression weighting coefficient was much larger for SR than for REPAIR mean, suggesting that SR contributed more to explaining the variance in Pre_S fluency ratings than did REPAIR mean. The scatter plots of the correlations are displayed in Figure 7.1.

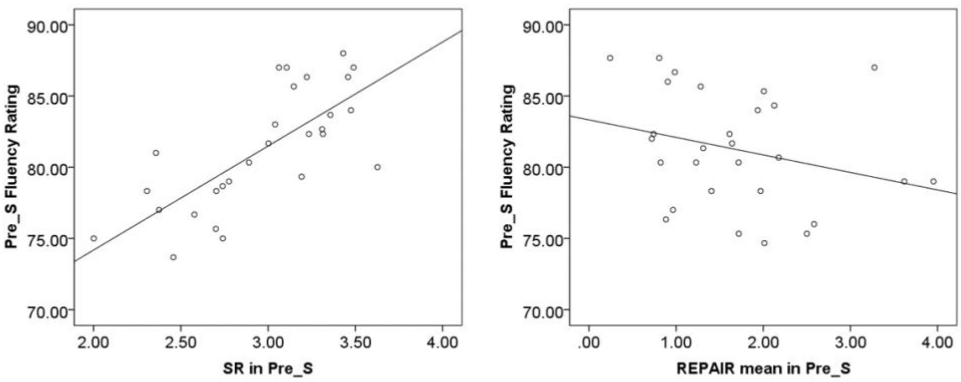


Figure 7.1 Scatter plots of correlations between SR and fluency ratings (left panel) and between REPAIR mean and fluency ratings (right panel) in the Pre_S task

7.1.2.2 Predicting role of utterance fluency indicators in Pre_F

Stepwise multiple linear regression analysis was conducted using the fluency rating in the Pre_F task (the pre-training task under conditions of high input rate) as the dependent variable and utterance fluency indicators as listed in Table 7.2 in the Pre_F SI performance as the independent variables. The stepwise multiple linear regression model was statistically significant [$F(2,27) = 21.763$, $p < 0.001$, $R^2 = 0.635$, *adjusted* $R^2 = 0.606$]. Results of the model are presented in Table 7.4.

Table 7.4 Coefficients of the regression model for Pre_F fluency rating

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	58.430	3.618		16.150	.000		
Pre_F_SR	8.554	1.296	.863	6.597	.000	.854	1.171
Pre_F_REPAIRmean	-1.279	.508	-.329	-2.516	.019	.854	1.171

The independent variables contributing to predicting the dependent variable were SR ($\beta = 0.863$, $t = 6.597$, $p < 0.001$) and REPAIR mean ($\beta = -0.329$, $t = -2.516$, $p = 0.019$

< 0.05). The remaining predictors were removed from the model. The results of the analysis showed that the variable that explained the greatest amount of variance of the dependent variable was SR, accounting for 54.3% of the variance of the dependent variable alone, and its correlation with the dependent variable was significantly positive with a coefficient of 8.554. The second variable added into the model was REPAIR mean, increasing the explained variance to 63.5%, and its correlation with the dependent variable was significantly negative with a coefficient of -1.279. The absolute value of the standardised regression weighting coefficient was larger for SR than for REPAIR mean, suggesting that SR contributed more to explaining the variance in Pre_F fluency ratings than did REPAIR mean. The scatter plots of the correlations are displayed in Figure 7.2.

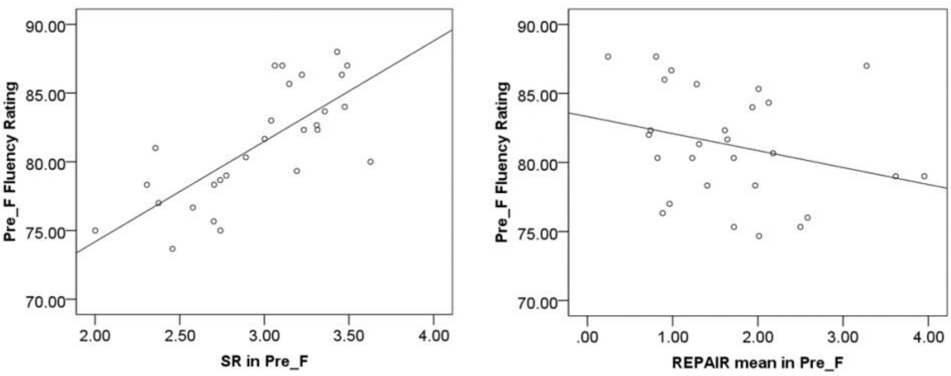


Figure 7.2 Scatter plots of correlations between SR and fluency ratings (left panel) and REPAIR mean and fluency ratings (right panel) in the Pre_F task

7.1.2.3 Predicting role of utterance fluency indicators in Post_S

Stepwise multiple linear regression analysis was conducted using the fluency rating in the Post_S task (the post-training task under conditions of low input rate) as the dependent variable and utterance fluency indicators as listed in Table 7.2 in the Post_S

SI performance as the independent variables. The stepwise multiple linear regression model was statistically significant [$F(2,27) = 16.246$, $p < 0.001$, $R^2 = 0.565$, *adjusted* $R^2 = 0.530$]. Results of the model are presented in Table 7.5.

Table 7.5 Coefficients of the regression model for Post_S fluency rating

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	64.440	3.381		19.061	.000		
Pos_S_SR	6.557	1.166	.760	5.626	.000	.954	1.048
Pos_S_FPmean	-1.047	.497	-.284	-2.105	.045	.954	1.048

The independent variables contributing to predicting the dependent variable were SR ($\beta = 0.760$, $t = 5.626$, $p < 0.001$) and FP mean ($\beta = -0.284$, $t = -2.105$, $p = 0.045 < 0.05$). The remaining predictors were removed from the model. The results of the analysis showed that the variable that explained the greatest amount of variance was SR, accounting for 48.8% of the variance of the dependent variable alone, and its correlation with the dependent variable was significantly positive with a coefficient of 6.557. The second variable added into the model was FP mean, increasing the explained variance to 56.5%, and its correlation with the dependent variable was significantly negative with a coefficient of -1.047. The absolute value of the standardised regression weighting coefficient was larger for SR than for FP mean, suggesting that SR contributed more to explaining the variance in Post_S fluency ratings than did FP mean. The scatter plots of the correlations are displayed in Figure 7.3.

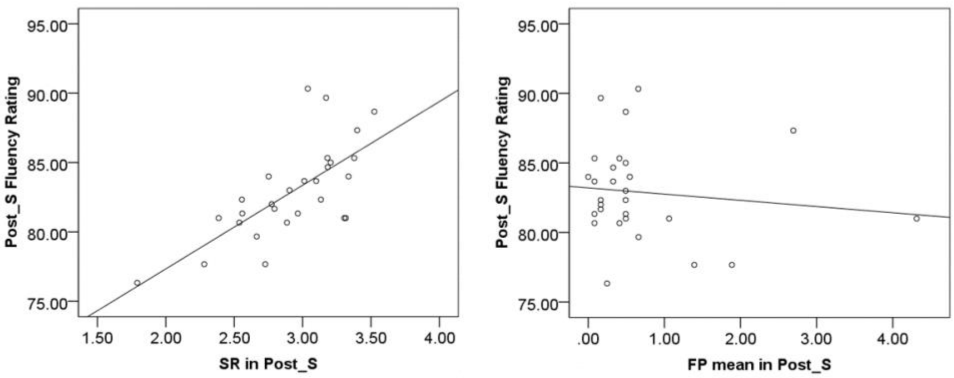


Figure 7.3 Scatter plots of correlations between SR and fluency ratings (left panel) and FP mean and fluency ratings (right panel) in the Post_S task

7.1.2.4 Predicting role of utterance fluency indicators in Post_F

Stepwise multiple linear regression analysis was conducted using the fluency rating in the Post_F task (the post-training task under conditions of high input rate) as the dependent variable and utterance fluency indicators as listed in Table 7.2 in the Post_F SI performance as the independent variables. The stepwise multiple linear regression model was statistically significant [$F(2,27) = 28.470$, $p < 0.001$, $R_2 = 0.695$, *adjusted* $R_2 = 0.670$]. Results of the model are presented in Table 7.6.

Table 7.6 Coefficients of the regression model for Post_F fluency rating

Model	Unstandardized Coefficients		Standardized Coefficients		Sig.	Collinearity Statistics	
	B	Std. Error	Beta	t		Tolerance	VIF
(Constant)	61.700	3.003		20.548	.000		
Pos_F_SR	7.101	.982	.809	7.228	.000	.973	1.028
Pos_F_FP mean	-.875	.263	-.372	-3.321	.003	.973	1.028

The independent variables contributing to predicting the dependent variable were SR ($\beta = 0.809$, $t = 7.228$, $p < 0.001$) and FP mean ($\beta = -0.372$, $t = -3.321$, $p = 0.003 < 0.05$). The remaining predictors were removed from the model. Results of the analysis showed that the variable that explained the greatest amount of variance was SR, accounting for 56.0% of the variance of the dependent variable alone, and its correlation with the dependent variable was significantly positive with a coefficient of 7.101. The second variable added into the model was FP mean, increasing the explained variance to 69.5%, and its correlation with the dependent variable was significantly negative with a coefficient of -0.875. The absolute value of the standardised regression weighting coefficient was larger for SR than for FP mean, suggesting that SR contributed more to explaining the variance in Post_F fluency ratings than did FP mean. The scatter plots of the correlations are displayed in Figure 7.4.

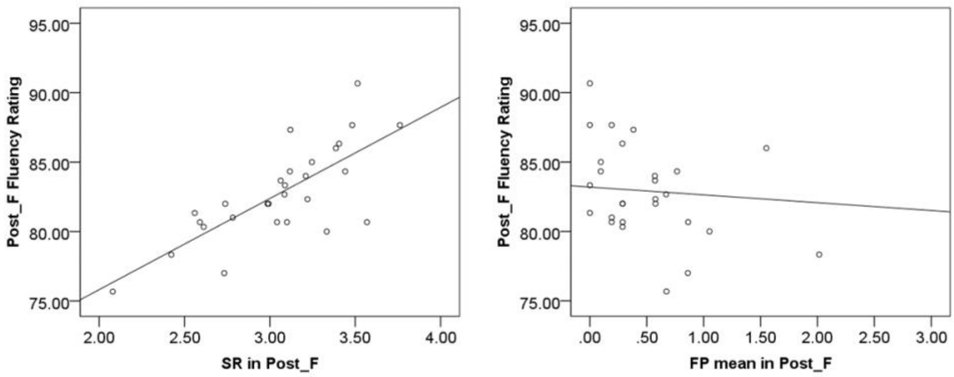


Figure 7.4 Scatter plots of correlations between SR and fluency ratings (left panel) and FP mean and fluency ratings (right panel) in the Post_F task

7.1.2.5 Diagnosis of the regression models

The residuals were diagnosed to check for multicollinearity, normality and homoscedasticity. The above multiple linear regression models are diagnosed below.

The VIF values of the independent variables of the four regression models were all smaller than 2, indicating that the correlation between them was comparatively weak.

There did not exist multicollinearity and the above results of the regression models were reliable. A PP diagram was used to investigate the normality of the residuals. As shown in Figure 7.5, most plots were scattered near the diagonals, indicating that the residuals of the linear regression models obeyed normal distribution, further verifying the accuracy of the results of the linear regression models.

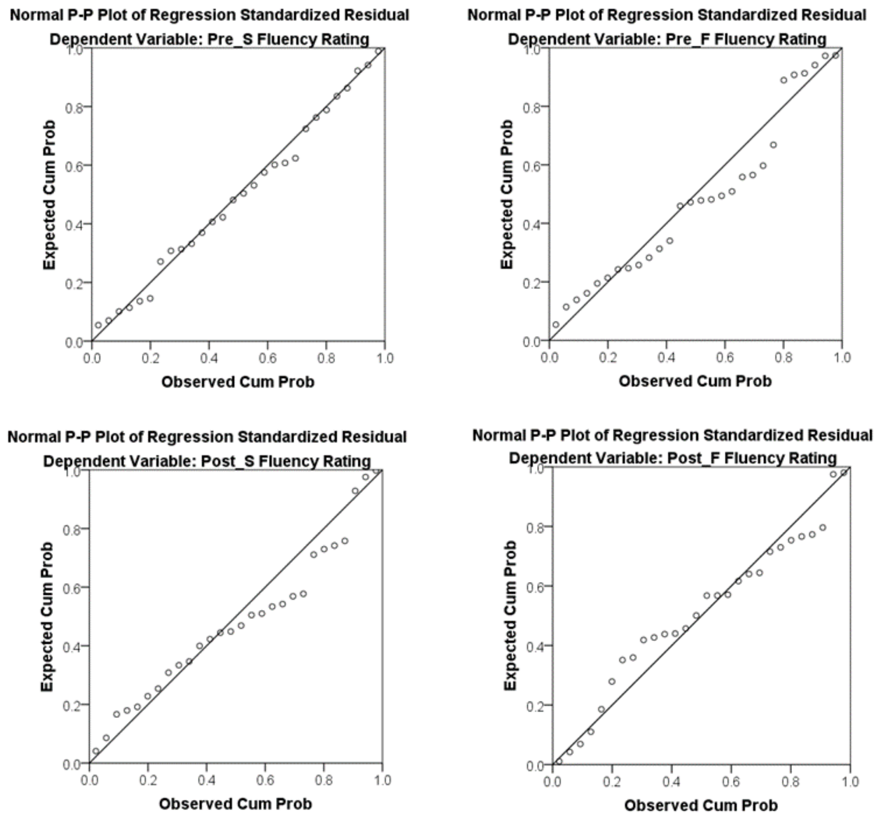


Figure 7.5 Normal P-P plots of models for speed fluency indicators

Results of the heteroscedasticity diagnosis are shown in Figure 7.6. The values of standardised residuals were all very small, mostly falling between -2 and 2. The residuals were distributed randomly, meaning there was no heteroscedasticity and further verifying the accuracy of the results of the linear regression models.

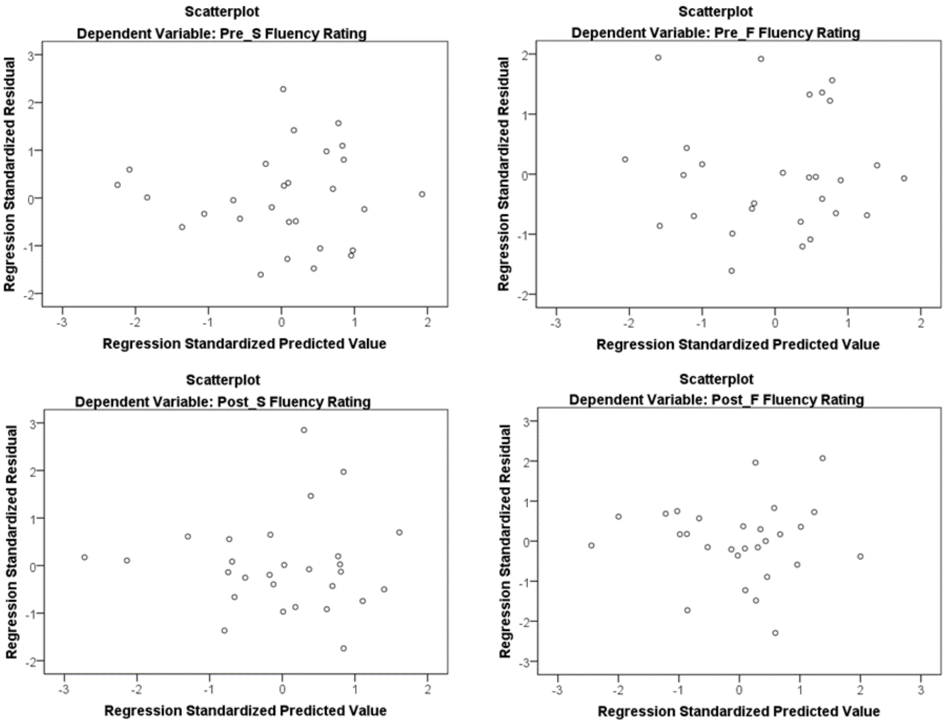


Figure 7.6 Scatterplots of models for speed fluency indicators

7.2 Summary and discussion of results

7.2.1 Summary of results of the regression analyses

Multiple linear regression analyses with utterance fluency indicators as independent variables and fluency ratings as dependent variables yielded models with two significant predictors for each of the four interpreting tasks (Pre_S, Pre_F, Post_S and Post_F). The results of the regression analyses are summarised in Table 7.7, listing the predicting power of the models (R^2) and the type of correlations (positive or negative) between independent variables that were entered into the models and dependent variables.

Table 7.7 Summary of the results of stepwise multiple linear regression analyses

	Predictors	Pre_S	Pre_F	Post_S	Post_F
1	SR	51.4%a↑	54.3%a↑	48.8%a↑	56.0%a↑
2	REPAIR mean	13.1%b↓	9.2%b↓		
	FP mean			7.7%b↓	13.5%b↓
	Model R^2	64.5%	63.5%	56.5%	69.5%

Notes: a. R^2 of the model with predictor SR
b. increase of R^2 with the second predictor entered
↑ positive correlation with the dependent variable
↓ negative correlation with the dependent variable

As summarised in the above table, the regression models with objectively measured utterance fluency indicators as independent variables accounted for 56.5-69.5% of the variance of fluency ratings of the trainee interpreters' SI performance. Under all four conditions, speech rate (SR) made a significant contribution to explaining the variance in fluency ratings and correlated positively with the dependent variables. SR made a major contribution in predicting perceived fluency, explaining around 48.8-56.0% variance of dependent variables alone. The mean number of REPAIRs (REPAIR mean) and mean the number of filled pauses (FP mean) were the second most significant predictors for fluency ratings in pre-training tasks and post-training tasks respectively, both correlating negatively with perceived fluency, but the increase in the explained variance was not as remarkable as with SR, ranging from 7.7% to 13.5% for the four SI tasks.

The regression models with SR and REPAIR mean as predictors accounted for a similar amount of variance of fluency ratings under low and high input rate conditions in the pre-training tasks, 64.5% and 63.5% respectively. In the post-training tasks, the regression model with SR and FP mean as predictors under the low input rate condition accounted for an obviously lower amount of the variance of the fluency ratings compared to that under the high input rate condition, 56.5% vs. 69.5% respectively. To be more specific, the predicting power of SR was much lower under the low input rate condition than under the high input rate condition in the post-training tasks (48.8% vs. 56.0%). The same pattern was observed for the predicting power of FP mean in the post-training tasks (7.7% vs. 13.5%).

A longitudinal comparison shows that the predicting power of the regression models with utterance fluency indicators as independent variables was obviously lower in the post-training tasks than in the pre-training tasks under low input rate conditions, 56.5% vs. 64.5% respectively. However, the trend seemed to reverse under high input rate conditions, with slightly more variance of fluency ratings explained in the post-training tasks than in the pre-training tasks.

7.2.2 Discussion

The current research, with 28 trainee interpreters at the MA level as participants, has confirmed that objectively measured indicators of utterance fluency are closely related to and can predict ratings of perceived fluency to a large extent, which sheds light on automatic assessment of interpreting. The results of the analyses were generally in line with findings of previous studies on the relationship between utterance and perceived fluency in interpreting, though differences existed in terms of the best predictor of perceived fluency and the predicting power of utterance fluency measures. Fluency ratings were found positively correlated to speed fluency measures and negatively correlated to breakdown fluency measures, with the phonation time ratio and speech rate best predicting to which fluency group professional interpreters belonged in Han's (2015a) study. Judged fluency was best predicted by the effective speech rate in Yu and van Heuven's (2017) exploration of the correlation between acoustic measures and judged fluency of trainee interpreters' L1-L2 consecutive interpreting performance. It should be noted that the participants in the two studies were professional interpreters and trainee interpreters, seven undergraduates and five MA students for Han (2015a) and Yu and van Heuven (2017) respectively, and the interpreting mode was consecutive in Yu and van Heuven's (2017) study. The differences in the level of interpreting expertise and the mode of interpreting could have led to differences in the findings. Moreover, in the current research, the speech rate, though it made the most significant contribution to explaining the variance of perceived fluency, was correlated to some of the other utterance fluency indicators (phonation time ratio and mean length of silent pauses). Thus, utterance fluency should be perceived as an integrated and dynamic construct when predicting perceived fluency.

Speed fluency variables seemed to be better predictors of perceived fluency compared to other dimensions of fluency. SR made the most significant and major contribution to explaining the variance of perceived fluency, and speed fluency variables were the main type of utterance fluency indicators (SR, PTR and MLR) that correlated significantly with perceived fluency in all four tasks. This is generally in line with previous findings regarding speech of second language learners (Bosker et al., 2012; Cucchiarini, Strik, & Boves, 2000; Kormos & Dénes, 2004), though the specific task performed and the speech material might affect the adequacy of the employed variables (Cucchiarini et al., 2002). Cucchiarini et al. (2000) pointed out that one possible reason that speech rate was a good predictor of perceived fluency might be that it incorporates speed and pause measures.

Moreover, the predicting power of disfluencies, i.e. the mean number of REPAIRs (repairs, repetitions and false starts) or the mean number of filled pauses, for perceived fluency is confirmed in this research, which has implications for interpreting pedagogy. The strength of correlations between metrics of repair fluency and fluency ratings were very weak in the study conducted by Han (2015a). In the study on L1-L2 consecutive interpreting performance by Yu and van Heuven (2017), the number of filled pauses contributed to explaining the variance of fluency ratings but the number of other disfluencies (in that study: repetitions, restarts, false starts and corrections) did not enter the regression models explaining the variance of fluency ratings, though they were significantly correlated to fluency ratings. This finding echoes that of Pinget et al. (2014) that the predictive power of repair fluency metrics (in their research: number of corrections and repetitions) for second language fluency assessment were non-negligible, in contrast to findings of previous research that disfluency phenomena did not influence or contributed little to perceived fluency (Bosker et al., 2012; Cucchiarini et al., 2002; Kormos & Dénes, 2004).

Comparisons of the predicting power of utterance fluency indicators for perceived fluency under conditions of low and high input rate in the post-training tasks indicated an obviously lower predicting power of utterance fluency under the low input rate condition, with less variance explained for each of the predictors (SR and FP mean). Such comparisons for the pre-training tasks did not lead to much difference. Longitudinal comparisons of the regression models for perceived fluency between the pre-training and post-training tasks showed that the predicting power of utterance fluency indicators for perceived fluency under low input rate conditions was much lower after training, while the trend was reversed under conditions of high input rate, though the difference was not substantial. One possible reason is that rating of fluency was influenced by the overall interpreting performance to some extent. Trainee interpreters had developed a certain degree of interpreting expertise after a period of intensive SI training, and were thus able to fulfil the interpreting task under the low input rate condition in a much more satisfactory way. This was proved by ratings of the overall SI performance, which indicated that the overall performance in the Post_S task received the highest ratings on average among the four tasks. Compared to the pre-training task, participants were also able to respond to the challenge of higher input rate of the source speech better after training, though their performance was not as satisfactory as for the low input rate speech. According to the results of the self-rating questionnaire (Appendix 7), participants found the post-training SI task with high input rate (Post_F) was not easy, though they felt that they completed the task better compared to the correspondent pre-training one.

Chapter 8 Conclusion

This chapter recapitulates the main findings of this research, explains its implications and significance, acknowledges several limitations, and points out directions for future research.

8.1 Summary of the main findings

This research followed 28 trainee interpreters at the MA level over an SI training period of thirteen weeks and offered a multidimensional exploration of fluency in their L2 (English)-L1 (Chinese) SI performance from the perspectives of cognitive, utterance and perceived fluency. Following a 2 (training) * 2 (input rate) factorial design, the underlying interconnections between the three domains of fluency in trainee interpreters' SI performance were investigated.

The main research questions explored in the current research included the role of cognitive fluency in the development of utterance fluency in the SI output of trainee interpreters under conditions of low and high cognitive load, how trainee interpreters' cognitive fluency measures, SI training and the input rate of source speeches influence their utterance fluency in the interpreting output, and the predicting power of objectively measured indicators of utterance fluency in trainee interpreters' SI output for perceived fluency as assessed by human raters. The main findings are summarised below in response to the three main research questions as listed in section 1.3.

8.1.1 The role of cognitive fluency in the development of utterance fluency

In terms of the relationship between cognitive fluency and the development of utterance fluency in the SI (L2-L1) output of trainee interpreters, results of a series of multiple linear regression analyses with cognitive fluency measures as independent variables and each of the measures of utterance fluency development (partialling out measures of utterance fluency in the pre-test from those in the post-test) as the dependent variables are summarised below.

1) Cognitive fluency measures could predict to a large extent the development of utterance fluency in trainee interpreters' SI output over an SI training period of thirteen weeks. The roles of individual cognitive fluency measures are summarised below.

a) The role of lexical access and lexical retrieval in utterance fluency development should be distinguished and treated separately in the investigation into the processes of SI. The efficiency of lexical access in the source language (LA EN) had a significantly positive influence on gains in the speech rate (SR) when the input rate was high and a positive influence on the mean number of silent pauses (SP mean) when the input rate was low. The efficiency of lexical access in the target language (LA CH) had a significantly negative influence on the gains in SR and mean length of run (MLR) and a positive influence on the change in mean duration of repairs, repetitions and false starts (REPAIR length) under conditions of high input rate, while it did not significantly influence the development of utterance fluency in the interpreting output under low input rate conditions.

b) The efficiency of lexical retrieval had a limited predicting power for the change in SP mean when the cognitive load was comparatively low and did not significantly influence the development of utterance fluency under conditions of high cognitive load.

c) The efficiency of linguistic attention control in the target language (AC CH) could predict the development in multiple indicators of utterance fluency while the efficiency of linguistic attention control in the source language (AC EN) did not influence the development of utterance fluency under conditions of either low or high input rate. AC CH had a significantly negative influence on the change in articulation rate (AR) and mean duration of silent pauses (SP length) and a positive influence on gains in MLR and phonation time ratio (PTR) under high input rate conditions, but it only had a significantly positive influence on the change in mean duration of filled pauses (FP mean) under conditions of low input rate.

d) Working memory capacity in the target language (SS CH), elicited from the speaking span task, had a significant influence on utterance fluency development in L2-L1 SI performance while working memory capacity in the source language (SS EN) did not. Working memory capacity only correlated significantly with gains in MLR and PTR under conditions of high cognitive load and the contribution of working memory capacity was independent. The finding confirms the domain-specific view of working memory and the proposition that the effects of working memory in interpreting are manifest in capacity-demanding tasks.

2) The efficiency of cognitive processes involved in the target language production stage, rather than that of processes involved in the source language comprehension

stage, had a main significant influence on the development of utterance fluency in SI performance.

3) The influence of cognitive fluency measures on SI utterance fluency development was evidently stronger under conditions of high cognitive load. Cognitive fluency measures had a significant influence on gains in all three dimensions of utterance fluency in trainee interpreters' SI output under conditions of high input rate, while they were only significantly related to gains in breakdown fluency measures under conditions of low input rate.

4) Measures of lexical retrieval and working memory capacity generally increased the explanatory power of cognitive fluency measures for the utterance fluency development of trainee interpreters' SI output. Lexical retrieval made an independent contribution to the change in SP mean in the SI output under conditions of low cognitive load while it did not play a significant role in the development of SI utterance fluency under high cognitive load. Working memory capacity made an independent contribution to gains in SR, MLR and PTR as indicators of trainee interpreters' SI utterance fluency under conditions of high cognitive load, but its contribution was not significant under conditions of low cognitive load.

8.1.2 The influence of cognitive fluency, SI training and input rate on utterance fluency

As for the influence of cognitive fluency, SI training and input rate on utterance fluency in trainee interpreters' output, a series of repeated measures ANOVA analyses were conducted, with SI training and input rate as within-subjects variables and each of the cognitive fluency measures as the between-subjects variable, to compare measures of utterance fluency between the pre-training and post-training SI performance, between conditions of low and high cognitive load, and between groups of low and high cognitive fluency efficiency. Results of the analyses are generalised below.

1) Interaction effects between cognitive fluency, training and input rate

The interplay between cognitive fluency, SI training and input rate existed for certain utterance fluency indicators, but the interaction was not common. Two-way interaction effects between training and input rate were found only for the mean number of REPAIRs (REPAIR mean). Further analyses of the simple effects indicated that training effects were only significant under conditions of high input rate where there were fewer REPAIRs per minute. The effects of input rate for REPAIR mean were only significant in the pre-test, which indicated that trainee interpreters could cope with high cognitive load better after training. Two-way interaction effects were also found between input rate and several of the cognitive fluency measures, including those of lexical access, lexical retrieval, linguistic attention control and working memory (LA EN, LA CH, LR, AC CH and SS EN). Further analyses of the simple effects showed that the main effects of input rate were significantly prevalent for nearly all cognitive fluency groups while only one cognitive fluency measure, i.e. working memory capacity in the source language (SS EN), had a significant influence on SR, which implied that the speech rate of the high efficiency SS EN group was significantly higher under high input rate conditions.

2) Influence of cognitive fluency

Cognitive fluency measures in the target language version generally had a significant influence on one or more indicators of utterance fluency in trainee interpreters' SI output and the role of cognitive fluency was generally independent. Moreover, it was noted that a higher efficiency of cognitive fluency did not necessarily contribute to a more fluent SI output.

Findings of the influence of individual cognitive fluency measures are summarised below.

a) The main effects of lexical access in the target language (LA CH) on the mean duration of filled pauses and the mean duration of REPAIRs (FP length and REPAIR length) were significant, with longer FP length and REPAIR length for the high efficiency LA CH group than those of the low LA CH group under conditions of both low and high cognitive load and in both the pre-test and the post-test.

b) The main effects of lexical retrieval (LR) on speech rate (SR), mean length of run (MLR) and phonation time ratio (PTR) were significant and the effect of LR on the mean duration of silent pauses (SP length) approached significance. Trainee interpreters with high LR efficiency might produce more fluent interpreting output, indicated by faster SR, longer MLR, higher PTR and shorter silent pausing time.

c) Linguistic attention control in the target language (AC CH) had a significant influence on the mean duration of REPAIRs (REPAIR length) and the high efficiency AC CH group exhibited longer duration of REPAIRs in their SI output under conditions of both low and high input rate and in both the pre-test and the post-test.

d) Working memory capacity in the target language (SS CH) influenced the mean number of REPAIRs (REPAIR mean) significantly, with more REPAIRs per minute in the SI output of the high SS CH group than in that of the low SS CH group under conditions of both low and high input rates in both the pre-test and the post-test.

3) Influence of SI training

The SI output after training was generally more fluent than that before training. The main effects of SI training were significant for speech rate (SR), articulation rate (AR) and mean length of silent pauses (SP length), with higher SR and AR and longer mean duration of silent pauses after training. The fact that there were fewer REPAIRs per minute under high input rate conditions after training provided further support for this finding.

4) Influence of input rate

The main effects of input rate were strongly significant for speed fluency indicators, including SR, AR and MLR, and were overall significant for the breakdown fluency indicators of the mean number of silent pauses and filled pauses (SP mean and FP mean) in the SI output, with faster SR and AR, longer MLR and fewer silent pauses but more filled pauses under high input rate conditions. In terms of indicators of repair fluency, input rate had a significant or near-significant impact on the mean duration of REPAIRs (REPAIR length), except for the LR and AC EN groups, with a shorter mean duration of REPAIRs under high input rate conditions.

In general, utterance fluency in trainee interpreters' SI performance was generally enhanced under conditions of high input rate, providing new evidence for the view that interpreters' SI output might be more fluent under high input rate conditions.

8.1.3 The predicting power of utterance fluency for perceived fluency

Regarding the relationship between utterance fluency and perceived fluency in SI, correlation analyses and multiple linear regression analyses were performed with objective measures of utterance fluency as independent variables and perceived fluency as rated by human assessors as the dependent variable. The analyses were conducted for conditions of low and high input rates in the pre-training and post-training tasks separately. The results of the analyses are summarised below.

1) The objectively measured utterance fluency indicators could account for a majority (56.5-69.5%) of the variance of perceived fluency of trainee interpreters' SI performance under conditions of low and high cognitive load and in both the pre-training and post-training tasks.

2) Speech rate (SR) had the strongest correlation with perceived fluency as rated by assessors and made a major positive contribution to predicting fluency ratings, explaining about half of the variance of fluency ratings alone.

3) The mean number of REPAIRs (repairs, repetitions and false starts) and the mean number of filled pauses (FP mean) were the second most significant predictors for perceived fluency in the pre-training and post-training SI tasks respectively and were negatively correlated with fluency ratings. The predicting power of disfluencies in general for perceived fluency in interpreting was confirmed and non-negligible.

4) In general, speed fluency variables seemed to be better predictors of perceived fluency in SI performance, compared to dimensions of breakdown and repair fluency.

8.2 Implications and significance

This research offers an interdisciplinary exploration of fluency in trainee interpreters' SI, integrating interpreting studies and studies in psycholinguistics and bilingual language production and broadening the perspective of cognitive fluency research. With reference to studies on fluency in bilingual language production, this research taps into multi-dimensions of fluency in SI, covering not only measures of utterance fluency and perceived fluency, but also the role of cognitive fluency. Moreover, it pays attention not only to measures of utterance fluency per se but also to its development and growth, providing references for future related studies on fluency in interpreting. The exploration of the relationships of the three domains of fluency in interpreting offers insights into the understanding of constructs of fluency in interpreting. The results of this research may enrich the existing knowledge of fluency in interpreting from multiple perspectives, contribute to the development of the theoretical framework of cognitive fluency and fluency in interpreting, and have implications for interpreting pedagogy.

Theoretically, the findings of this research provide evidence for the effectiveness of including lexical retrieval and working memory capacity in the exploration of cognitive fluency in interpreting, which has implications for the theoretical framework of cognitive fluency in interpreting and bilingual language production. Exploration of the role and constructs of cognitive fluency enhances the understanding of the information processing mechanism of SI and the construction of relevant theoretical models of fluency in interpreting.

Methodologically, this research adopts a 2*2 factorial design, with input rate and interpreting training as two factors and extraneous variables well controlled, providing references for related studies on fluency in interpreting. Measures of cognitive fluency tasks adopted the CV (coefficient of variance) index to measure the efficiency of the cognitive processes instead of focusing on the reaction time. Moreover, the composite strict speaking span was adopted in this research for the measurement of working memory capacity. It takes processing accuracy, processing efficiency and storage ability into consideration and has been proved to be an effective index of working memory capacity. These methodological considerations have implications for future related studies.

Pedagogically, the finding that trainee interpreters' SI output (L2-L1) tended to be more fluent under high input rate conditions after a period of training sheds light on interpreting training. It calls for the need to adjust and diversify training material based on the level of expertise development of trainee interpreters, even at the initial stage of interpreter training. Identifying the influence of cognitive fluency on utterance fluency and overall SI performance has implications for the interpreting aptitude test. The close correlations between indicators of utterance fluency and perceived fluency shed light on the automatic assessment of interpreting quality. The predicting power of disfluencies in the interpreting output for perceived fluency is further confirmed, which has implications for interpreting pedagogy and draws more attention to the importance of the control of disfluencies in interpreting production. Moreover, explorations of the contributing factors of utterance fluency provide references for possible improvement of training methods to promote the development of fluency in SI as well as interpreter competence.

8.3 Limitations and future research

8.3.1 Limitations of this research

Several limitations of this research are acknowledged. The size and nature of the sample constitutes the first limitation of this research. Twenty-eight participants produced 112 samples of SI performance under four different conditions. All participants were unbalanced trainee interpreters at the MA level, and most were Chinese native speakers, constituting a comparatively homogeneous sample. The number of raters assessing the SI performance was limited, though the agreement among raters was strong. A large and heterogeneous sample would contribute to verifying the findings of the current research.

The second limitation is the research scope. The research only explores L2 (English) to L1 (Chinese) SI instead of bi-directional performance. Moreover, though efforts were made to present a full picture of fluency in interpreting, the research explored the influence of training and cognitive load on fluency in SI but did not include other variables like the level of interpreting expertise, which are important factors of fluency in interpreting. One important reason for the current choice was to keep the experimental design within a manageable scope and to avoid overloading the participants and affecting the internal validity.

The third limitation of this research is the ecological validity. The experimental setting has its strength in the control of extraneous variables, for instance, the comparability of source speeches, but the experimental nature of the current research means that its ecological validity is compromised to some extent, though efforts were made to simulate real-life SI settings. Another limitation is that the training period of thirteen weeks is comparatively short, so the training effects might not be as evident as they would be over a longer training period.

8.3.2 Suggestions for future research

Future studies may extend the research scope by tapping into fluency in L1-L2 SI and including more language pairs to verify the findings of this research, in particular relationships between different domains of fluency. The inclusion of a heterogeneous sample of participants, for instance participants of different levels of interpreting expertise, and the tracking of a longer period of interpreting training might lead to more multifaceted findings.

In terms of cognitive fluency, its constructs and their specific roles in interpreting should be further explored in future studies. Taking attention control as an example, the current research focuses more on the shifting function of attention control using linguistic stimuli in the behavioural experiment. It is worth further exploring whether other functions of attention control play similar roles in the interpreting process and contribute to fluency in the interpreting output in the same way. The addition of lexical retrieval and working memory capacity in the framework of cognitive fluency in interpreting is proved to be effective in this research. Further investigation into more cognitive tasks might be a direction for future related work to facilitate the theoretical development of fluency in interpreting and bilingual language production.

As for utterance fluency, automatic measurement of utterance fluency indicators with speech analysis tools should be adopted to facilitate fluency related studies in the future. Exploration of a more efficient and scientific way to measure utterance fluency should be encouraged. In terms of perceived fluency, the current research operationalised perceived fluency as ratings of assessors. Future research should also pay attention to user perception of interpreting performance and may include raters or listeners of different background to enhance the reliability of relevant findings.

Appendices

Appendix 1 Sheet of participants information

Code: ____________

Personal Information:

Name:	______________________________
Gender:	Female ☐ Male ☐
Birth Date (dd-mm-yyyy):	______________________________
First Language:	______________________________
Second Language:	______________________________

Education/working Background:

Interpreting training:	______________________________
Interpreting experience:	______________________________
Level of education:	______________________________

English competence:

IELTS / TOEFL (Circle)	Listening ______ Speaking ________
Overall Score _______	Reading _______ Writing ________
Current English comprehension (1-5):	[Self-rated] ______________________

______________________________ ________________ ______________________

(*Participant*) (*Date*) (*Signature*)

Appendix 2 Wordlist of the animacy judgment task

English word list:

Practice block:

Animate: rat, sheep, son, teacher, tiger
Inanimate: stamp, street, tape, television, window

Experimental block:

Animate: actor, adult, ant, bear, bee, bird, boy, bride, brother, cat, child, cow, dancer, daughter, dentist, doctor, dog, duck, farmer, father, girl, goat, horse, human, husband, judge, king, lady, monkey, mother

Inanimate: basket, belt, bench, bicycle, blanket, board, boat, book, building, car, chair, chimney, clothes, comb, desk, dictionary, door, fireplace, floor, garbage, ink, job, key, kitchen, knife, lamp, luggage, medal, newspaper, pants

Chinese word list:

Practice block:

Animate: 老鼠、绵羊、儿子、教师、老虎
Inanimate: 邮票、台阶、街道、磁带、电视

Experimental block:

Animate: 演员、成人、蚂蚁、狗熊、蜜蜂、小鸟、男孩、新娘、兄长、小猫、儿童、奶牛、舞者、女儿、牙医、医生、猎狗、鸭子、父亲、护士、女孩、山羊、秘书、丈夫、农民、法官、国王、女士、猴子、母亲

Inanimate: 裤子、丝带、篮子、板凳、自行车、地毯、木板、飞机、书本、建筑、汽车、椅子、烟囱、衣服、梳子、桌子、词典、道路、壁炉、地板、车库、墨水、工作、钥匙、厨房、刀具、台灯、行李、踏板、报纸

Appendix 3 Wordlist of the word translation task

Practice block:

stone, friend, fruit, summer, street, pen, pearl, thief, peach, pear

Experimental block:

Low-frequency stimulus words:

shoulder, coffee, music, house, book, river, heart, doctor, hospital, war, dress, road, body, wall, meat, car, watch, coat, shop, town, tree, skin, farm, army, end, feeling, form, nature, manner, method, influence, reason, result, chance, action, culture, language, promise, direction, difference, duty, truth, future, experience, science, memory, solution, century, change, possibility

High-frequency stimulus words:

lamb, rice, crown, violin, slave, cotton, flame, pirate, monkey, grape, lemon, pillow, turtle, bullet, bike, sleeve, spoon, signature, moustache, carrot, airplane, butterfly, umbrella, height, thread, sweat, silver, betrayal, climate, warmth, rhythm, whale, murderer, deed, luck, domain, inheritance, patience, honesty, coward, regret, innocence, conscience, treaty, curse, abuse, disadvantage, estimation, cruelty, despair

Appendix 4 Wordlist of the category judgment task

English version

Stimuli for “time”:

Words indicating “past”: *ago, past, yesterday, just now*

Words indicating “future”: *afterwards*, *future, tomorrow, soon*

Stimuli for “frequency”:

Words indicating low frequency: *rarely, occasionally, seldom, never*

Words indicating low frequency: *common, often, frequently, always*

Chinese Version

时间词:

表示“过去”的词：以前、 过去、 昨天、 刚才

表示“将来”的词：以后、 将来、 明天、 马上

频率词：

表示“频率低”的词：罕见、偶尔、 很少、从不

表示“频率高”的词：屡次、经常、频繁、总是

Appendix 5 Wordlist of the speaking span task

English version:

Practice Trial

Set 1: prevent, benefit

Set 2: address, brother, factory

Set 3: message, college, compare, holiday

Set 4: courage, village, suppose, kitchen, officer

Set 5: opinion, station, illness, replace, clothes, mention

Series 1

Set 1: ability, private

Set 2: example, produce, society

Set 3: surface, include, capital, another

Set 4: control, century, failure, reading, special

Set 5: medical, subject, develop, outside, chapter, meeting

Series 2

Set 1: meaning, natural

Set 2: concern, believe, present

Set 3: average, popular, develop, certain,

Set 4: country, problem, respect, account, process,

Set 5: attempt, project, feeling, second, foreign, reality

Series 3

Set 1: similar, history

Set 2: content, support, quality

Set 3: serious, product, measure, company

Set 4: science, service, patient, thought, success,

Set 5: several, provide, suggest, general, teacher, disease

Chinese version:

Practice Trial

Set 1: 阻止、受益

Set 2: 地址、兄弟、工厂

Set 3: 消息、大学、比较、假期

Set 4: 勇气、农村、假设、厨房、官员

Set 5: 主意、车站、疾病、代替、衣服、提到

Series 1

Set 1: 能力、隐私

Set 2: 举例、生产、社会

Set 3: 表面、包括、首都、其他

Set 4: 控制、实际、失败、阅读、特别

Set 5: 医疗、主题、发展、外面、章节、会议

Series 2

Set 1: 意义、自然

Set 2: 担心、相信、礼物

Set 3: 平均、流行、发展、确定

Set 4: 国家、尊重、问题、账户、过程

Set 5: 试图、项目、感觉、国外、举手、现实

Series 3

Set 1: 相似、历史

Set 2: 内容、支持、质量

Set 3: 严肃、产品、衡量、公司

Set 4: 科学、病人、思考、耐心、成功

Set 5: 一些、提供、词典、普通、老师、疾病

Appendix 6 Procedure description of the SI tasks

Description of the Procedure of the Simultaneous Interpreting Task

You will interpret two speeches from English to Chinese. Each speech lasts for about 12 minutes.

The two speeches are delivered by the same speaker but of different speech rates. Please try your best to interpret accurately and fluently and avoid being silent for too long time.

There is a warm-up session. Please adjust your headset, microphone and the volume of the source speech to make yourself comfortable. You may stop and ask questions any time during this session.

Before interpreting for each speech, please read the briefing note, with the information of the speaker, the background, summary, and glossary of the speech to get familiar with the context of the speech.

While interpreting, you will see the live video of the speech. The speeches are addressed to a large audience in an auditorium. Please imagine you are interpreting in real-life situations and the audience cannot understand the original. There is a short questionnaire after each SI task.

After interpreting for speech one, you will take a 10 minutes break before interpreting the second speech. The whole process of SI tasks will take about 40 minutes.

Appendix 7 Questionnaire for interpreters

Participant No. / Name:

Questionnaire of the SI task for the speech (Please circle your choice)

1. Please rate the speech rate of the original speech you just interpreted.

1 very slow 2 slow 3 medium 4 fast 5 very fast

2. Please rate the level of difficulty of the original speech you just interpreted.

1 very easy 2 easy 3 medium 4 difficult 5 very difficult

3. Please rate your overall SI performance.

1 very unsatisfactory 2 unsatisfactory 3 medium 4 satisfactory 5 very satisfactory

4. Please rate the level of accuracy in your SI output.

1 very inaccurate 2 inaccurate 3 medium 4 accurate 5 very accurate

5. Please rate the level of fluency in your SI output.

1 very disfluent 2 disfluent 3 medium 4 fluent 5 very fluent

6. Please rate the output rate of your delivery in the SI output.

1 very slow 2 slow 3 medium 4 fast 5 very fast

7. Please rate your overall control of silent and filled pauses in your SI output.

1 too bad 2 bad 3 medium 4 good 5 very good

8. Please rate your overall control of disfluencies (repairs, false starts and disfluent repetitions) in your SI output.

1 too bad 2 bad 3 medium 4 good 5 very good

Appendix 8 Briefing note of speech A in the pre-test

Prime Minister Lee Hsien Loong (Singapore)

National Day Rally speech

At the ITE College Central, 2013

Background

National Day Rally (NDR) is an annual address that the Prime Minister of Singapore makes to the entire nation, on the first or second Sunday after National Day on 9 August. At the rally, the Prime addresses the nation on its key challenges and announce major policy changes.

Speech Summary

The speaker talks about the challenges Singapore faces, like competition from technology, globalization, income inequalities. He states how Singaporean individuals, community and government should respond to these challenges in the new phase of development. Issues of housing and education are addressed.

Glossary

Our Singapore Conversation (OSC):
a national conversation initiative

Kampong Spirit:
A sense of social cohesion in a community where there are understanding and compromise among neighbors

Home Ownership
the housing program in Singapore

HDB flat
flats offered by HDB, the *Housing and Development Board* of Singapore

Appendix 9 Script of speech A in the pre-test

Good evening again. I hope you've enjoyed taking a look at the campus and meeting the students here at ITE College Central. I brought the Rally to ITE for a serious purpose to signal a change, to emphasize that this is not the usual NDR. Singapore is at a turning point. Tonight, I'll talk about the challenges which we face and what we must do to change to respond to these challenges in this new phase of our development and nation-building.

Last year, I spoke about the essential elements of our future, "Hope, Heart and Home". And since then, we've been holding the one, Our Singapore Conversation OSC, on building a better Singapore. The OSC took a fresh approach to engage Singaporeans. They had no pre-set agenda. They are a fully open discussion. And it elicited a very positive response. Nearly 50,000 people participated, and diverse groups.

But they also expressed what they wanted to see in Singapore. First of all, opportunities -opportunities to lead fulfilling lives, to realize one's potential. Secondly, purpose-coming together to build a better Singapore. Thirdly, assurance-assurance that our basic needs can be met, that we do not have to face life's uncertainties alone. Fourth, community spirit - closer community ties, stronger social cohesion, a warmer kampong spirit. And finally, and fifthly, trust - trust between the Government and people, trust among Singaporeans.

So the OSC has been a very meaningful exercise. We've listened to one another, we've created a firmer, shared basis to discuss and to plan our future.

To achieve our aspirations, we need to take into account the world around us. This is a time of rapid change and uncertainty. I’ve discussed these themes before many times. So tonight, I’d just like to briefly share with you a few striking facts, about technology, about globalization, about competition, income inequalities.

Technology is transforming our lives. Competition from technology, competition also from new emerging economies, China, India, Vietnam. China and India alone, one billion workers altogether. Every year, millions of new graduates entering the workforce. Just now in the Mandarin speech, I said seven million from China. If you added some more from India, it is ten million a year, all hungry, looking for work.

So we are seeing competition and we are seeing income inequality rising, the top zooming away, middle class stagnating. People with exceptional skills, globally in demand doing very well, not just IT or financial services, but even culture or sports.

So Singaporeans are affected by these global trends and feeling uncertain and anxious also. Because in Singapore too, technology and globalization are widening our income gaps. And in addition to that, we have domestic social stresses building. Population aging, society becoming more stratified, less mobile, children of successful Singaporeans more likely to do well. Children of lower income families, fewer of them rising than in previous generations. It's a reality. We acknowledge it. We have to do our best to do something about it. These trends are compounded by day-to-day problems: cost of living, public transport. You know them as well as I do. So Singaporeans sense correctly that the country is at a turning point. I understand your concerns. I promise you, you will not be facing these challenges alone because we are all in this together. We'll find a new way to thrive in this new environment.

We must make now a strategic shift in our approach to nation-building. Singapore has been built on three pillars: the individual, the community and the state. And each has played a role complementing each other. The individual, working hard, saving for himself and his family. The community, getting together to help different groups of people. And overall, the Government, creating the conditions for a vibrant economy and for good jobs, investing heavily in our people through education, through housing, through healthcare, but keeping state welfare low and targeted, stringent.

Today, the situation has changed. If we rely too heavily on the individual, their efforts alone will not be enough, especially among the vulnerable like the low-income families, like the elderly. And there are some things which individuals cannot do on their own. And there are other things which we can do much better together. So we must shift the balance. The community and the government will have to do more to support individuals. The community can and must take more initiative, organizing and mobilizing ourselves, solving problems, getting things done.

The government will also do more to support individuals and the community. What we used to do we will continue to do, to provide core public services, housing, education, healthcare. But at the same time, we will make three important shifts in our approach. First, we will do more to give every citizen a fair share in the nation’s success, raise the incomes and the wealth of the low-income Singaporeans, for example, through our housing program - home ownership.

Secondly, strengthen social safety nets. Assure people that whatever happens to you, you can get the essential social services that you need, especially healthcare.

Thirdly, do more to keep paths open upwards for all. To keep our society mobile, to bring every child to a good starting point and make sure that however, whichever family you are born to, whether you are privileged or not privileged, you are never shut out from the system, from opportunities, and especially through education. These are three strategic shifts. One, to level up people; two, to share the risks, to make sure that whatever happens in life, you will not be alone; and three, to keep our system open, mobile, so that if you have talent you can rise, if you work hard, you can get ahead. We will apply these shifts progressively to all our social policies.

And let me tonight talk about housing, and healthcare, and education, specifically at a little bit more detail, so that you understand what we are trying to do.

Singapore has succeeded because everyone has shared in the fruits of our progress. Incomes have risen across the board. The values of homes have appreciated. And even poor people are not poor by any international standard.

We have made…Let me summarize my points. We've made significant moves in recent years. Tonight, what I have announced is another significant move, but it is not the end of the story. We will monitor closely how well people can afford housing in Singapore. And over time, as it becomes necessary, we will do more to help the lower and the middle-income Singaporeans own their homes. We will always make sure that an HDB flat is always within reach, affordable and available to Singaporeans.

Besides housing, we'll also give Singaporeans more assurance over life uncertainties, especially healthcare. Working adults feel the pressure taking care of growing children, also looking after elderly parents. So, we will improve healthcare financing to give

Singaporeans more peace of mind.

The third major shift which we will make is to do more to keep paths upwards open to all, wide open to all. Keeping paths wide open has been a fundamental principle for Singapore for a very long time. It's how we've enhanced our human potential. How we've created hope for every Singaporean. And it's especially true in education.

I believe we can make every school a good school. And we have done a lot of that to ensure that every school provides a good education for the students. We give them the resources, we give them the good teachers, we emphasize values and we've made a lot of progress towards this goal.

But I am a realist. I accept that parents and students will always carefully choose which schools to go to. And I think it is good that parents compare and choose schools because it puts pressure on the schools to know that the parents are watching and that it makes a difference how they perform. But it's important that parents compare and contrast and choose on the right basis, not just examination grades, but also how well the schools are really educating their children.

We have an excellent education system. But our society is getting more stratified. Competition is intensifying amongst our students. And the focus, unfortunately I think, is too much on examination performance, not enough on learning.

Our new strategic direction will take us down a different road from the one that has brought us here so far. There is no turning back. I believe this is the right thing to do given the changes in Singapore, given the major shifts in the world. We proceed, but let me sound a word of caution: All this is not without risk.

Therefore, we have to tread carefully, beware the pitfalls. We will do more for the low income, but we cannot undermine self-reliance. We will increase healthcare spending, but we cannot encourage over-consumption and unnecessary treatments. We will make the education system broader and more open, but we cannot compromise academic standards and rigor. And finally, of course, all good things have to be paid for. Yes, for now, we can afford these measures from existing revenues. In the longer term, their costs will rise, especially healthcare costs.

Appendix 10 Briefing note of speech B in the pre-test

Prime Minister Lee Hsien Loong (Singapore)

National Day Rally speech

At the ITE College Central, 2016

Background

National Day Rally (NDR) is an annual address that the Prime Minister of Singapore makes to the entire nation, on the first or second Sunday after National Day on 9 August. At the rally, the Prime addresses the nation on its key challenges and announce major policy changes.

Speech Summary

The speaker talks about how Singaporeans respond to disruptive challenges in the globalized world. Challenges should be embraced, with new capabilities built and entrepreneurship promoted. He then states how Singapore can secure its place in the world. The relations between Singapore and the US and China are discussed.

Glossary

SAF: Singapore Armed Forces

ASEAN: Association of Southeast Asian Nations

Asian Infrastructure Investment Bank

Straits of Malacca

Appendix 11 Script of speech B in the pre-test

My fellow Singaporeans, good evening again. A lot has happened since my last National Day Rally. We had a good General Election and now we have the next generation of leaders firmly in place. We've been working hard to lay the foundation for the next 50 years.

But while we celebrate, we should continue to prepare for our future. Therefore tonight, I want to speak seriously about the challenges that we face. My subjects will not all be easy or fun ones. Some sensitive topics may make us, may even make us feel a little bit uncomfortable. But it is my responsibility to talk candidly about them and tell you honestly what lies ahead, and what we need to do to progress together.

I want to ask three questions. One, how do we progress together? Two, how do we secure our place in the world? And three, how do we ensure good politics for Singapore?

First question, how do we progress? We aspire to be among the leading cities in the world, New York, Shanghai, London, Tokyo, Sydney. One of the shining spots in the world, in human civilization, where people want to be. Because here, you can do great things. A good place to bring up children where our young continue to enjoy opportunities and a better life.

At the same time, we want Singapore to be our home, where we belong, where our families feel safe and secure, where we build a compassionate and inclusive society, leaving no one behind.

We know what we want, but how do we get there? Among all the economic issues we are dealing with, slower growth, helping people to upgrade, strengthening social safety nets. The defining challenge which we face in this era is disruption. Things are changing fast, old models are not working, new models are coming thick and fast. And we are having to adjust and to keep up, because of technology, globalization. And the disruption will happen over and over again relentlessly.

We can respond to disruptive change like this in two ways. We can close ourselves off, try to stop people from using the new technology. You must do things the way you have always done it. No, we impose restrictions, we protect their old ways, the taxi companies, and we force everybody to stick to that. But we will be left behind, and our commuters will lose out and our economy will suffer.

The other way is to embrace change. Let the disruption happen. You cannot stop it, but you can adjust it, you can adjust to it. Let the commuters enjoy better service, but help the incumbents and especially help the taxi drivers to adapt to the changes. That is what we are doing.

And as a transport and financial hub, if people are buying online all over Southeast Asia and Asia, we can be in that business. The goods shipped through Singapore, the financing has to be done somewhere. We can be a major player in this new logistics chain. And there will be new opportunities, data analytics, digital marketing, figuring out who is buying what, who is interested in which item, how to get the item to the person who wants it.

At the same time, we also need an overall strategy to find out, to be able to spot whether changes are coming from, to respond when things are disrupted and to keep our economy growing. So that our companies can be resilient, able to keep on finding new ways to do business, able to keep on employing Singaporeans in good jobs. And we have to prepare our workers to do good jobs, different jobs, new jobs, during their lives. First, to build new capabilities. Secondly, to promote entrepreneurship. And thirdly, to develop new skills amongst our people. First of all, we will help our companies to develop new capabilities. And one area where we can do this, which is promising, is digital because Singapore is well-connected. We've got fiber everywhere.

Besides digital, we also need to build capabilities in other sectors. We have some SMEs which can compete with the best in the world and we should help them to grow, to scale up. Beyond building capabilities, we should also promote entrepreneurship. Because engineering is one thing, but entrepreneurship, that dare to try something new, to venture, that is a different mindset. And entrepreneurs have that mindset, need that mindset and we need people like that in our society. They play an important role, not just because they are doing business for themselves, and doing well and creating jobs and prosperity and making it big. But also because they are resourceful, they are optimistic. They give the society confidence. A strong economy, therefore, needs capabilities. You need the entrepreneurs, but you also need a skilled workforce and if we give our workforce skills, we will enable them to hold better jobs, earn better pay.

I've talked about developing capabilities, encouraging entrepreneurship and developing skills. Earlier in my Chinese speech, I also discussed strengthening our safety nets. These are the ways our economy and our workers can thrive amidst disruption. This is how we can progress together and thrive in a competitive and dangerous world. And it is a dangerous world.

So my second question is, how do we secure our place in the world? To start off with, we have to defend ourselves and that is why we have the SAF. But as a small country, we also need a network of friends, friends in our neighborhood, and also friends among the big powers all over the world, even in faraway places.

We benefit from the dynamic and innovative US economy. We admire their warmth, their openness, their generosity. My visit was also a signal that the US values its friends and partners and appreciates Singapore's support for the role that America has played in the Asia-Pacific for more than 70 years since the war, spreading prosperity through trade and investments, maintaining security and stability, enabling all the countries in the region to thrive and to compete peacefully. And Singapore hopes that the US would keep on doing this even as China's influence grows.

Singapore hopes that China will develop and grow well too because an unstable and backward China will cause Asia great trouble, as happened in the 1950s and 1960s. Over the last 40 years, China has reformed and opened up under Deng Xiaoping and his successors. China has been stable and prosperous, and has, this has greatly benefited Asia and the world, including Singapore.

We are happy to see China grow strong and influential in a constructive and peaceful way. Upholding international law and the peaceful settlement of disputes, for a small country like Singapore, this is a vital interest. When we have disputes with other countries that's how we settle them.

Our second interest in the South China Sea is freedom of navigation. I'll explain why. You look at the map. Singapore is this little spot down here, a little red dot. We have two vital sea lanes of communication, two arteries, one through the South China Sea, one through the Straits of Malacca. So, it is really important to us that disputes in the South China Sea do not affect freedom of navigation or overflight by ships or aircraft.

Thirdly, Singapore needs a united and effective ASEAN. Our voice internationally, five million people, counts for not so much. But if ASEAN can get together, united, collectively, 600 million people, it can make itself heard quite a lot better, provided ASEAN is united.

Our relationship with China is much broader than the South China Sea. The friendship has lasted for decades. We are pursuing opportunities in infrastructure, in connectivity, financial services, urban planning, clean technology, working with China on One Belt, One Road, participating in the Asian Infrastructure Investment Bank. And we have many more opportunities to strengthen our friendship and cooperation with each other. That's how we do business.

I am sharing my concerns and plans with you because all of us have a role to play building Singapore together. But whom are we building Singapore for? It is not just for ourselves. It is for our children, our grandchildren. It's always been the Singapore story.

Recently, somebody asked me at a dialogue, 'If God appeared before you and offered you three wishes for Singapore, what would you ask for?' I paused. I was taken aback. I thought about it. I said, if I ask for material things, we will regret it, because after you've got it, you've consumed it, you've enjoyed it, you will not be satisfied, you will want more. But what I would like to have is that we be blessed with a 'divine discontent', always not quite satisfied with what we have, always driven to do better. At the same time, that we have the wisdom to count our blessings so that we know how precious Singapore is. And we know how to enjoy it and to protect it. And that if we have just these two wishes fulfilled; I think that is enough. Because then, then we can keep on keeping Singapore special and building something special in Singapore for many more years. And then, we can achieve happiness, prosperity, and progress for our nation. Thank you and goodnight.

Appendix 12 Briefing note of speech C in the post-test

Prime Minister Lee Hsien Loong (Singapore)

National Day Rally speech

At the ITE College Central, 2014

Background

National Day Rally (NDR) is an annual address that the Prime Minister of Singapore makes to the entire nation, on the first or second Sunday after National Day on 9 August. At the rally, the Prime addresses the nation on its key challenges and announce major policy changes.

Speech Summary

The speaker talks about how to give people full opportunities to achieve their potential, to provide Singaporeans more assurance in retirement, and to make the country a home for all ages.

Glossary

SG50: the 50 years of Singapore

ITE: Institute of Technical Education

The CPF scheme: Central Provident Fund, a social security savings plan that has provided many working Singaporeans with a sense of security and confidence for their retirement years

Home Ownership: the housing program in Singapore

Appendix 13 Script of speech C in the post-test

Singapore has come a long way. It is the work of generations, each standing on the shoulders of the one which came before, and it started with one special generation – the Pioneer Generation.

Our pioneers were ordinary people who worked together to do extraordinary things. They overcame difficult and dangerous situations to build a sovereign, independent country. They transformed Singapore from Third World to First and they always looked to the future and strove to give their children better lives than themselves. And this is why we are commemorating SG50 next year, to celebrate the spirit of our pioneers and to commit ourselves to their enduring values as we make our way forward.

Singapore is at a turning point. The world is in flux. Conflicts far away affect us.

Nearer home, we see tensions in the South China Sea. And the tensions are affecting sentiment in the region. It's affecting cooperation between countries. It's having an impact on confidence among businesses. It's even hardening attitudes among ordinary people towards other countries.

Singapore is changing too. There is a new generation with new aspirations. There is an aging population, which is creating new social needs. We have a better home, but we have the potential to do much much more.

So this year, I will focus on three topics – giving our people full opportunities to achieve their potential, providing Singaporeans more assurance in retirement, and making this Singapore, a home for all ages.

Singapore must always give our people full opportunities to achieve their potential. Our pioneers showed that we can do anything, provided we set our minds to it. And we must build on their legacy and continue to give every Singaporean the confidence to shoot for the stars. Education is an important part of this. And that's why every year I speak on different aspects of education.

And this year, I will focus on ITE and Poly students. Our ITEs and Polys are world-class. Foreign visitors are amazed by the facilities, better than many universities. Investors are impressed by the quality of the graduates, well-trained, can-do, productive. And our students themselves are great examples of resolve, strength and character.

They are right to aim high, we want to help them to create a brighter future for themselves by many routes, not just the academic route but also alternatively by getting good jobs, mastering deep skills, performing well and then getting relevant qualifications along the way, as they work, as they advance in their careers.

And we have to help individuals to progress and upgrade after they have graduated and started work in their careers to develop the structured career paths for them and then to implement this work and study path on a national scale. It's not easy. It involves multiple stakeholders, many government agencies: education, manpower, trade and industry.

But to succeed nationally, we need two strategic factors which will help everyone to achieve their potential. One, you must have economic growth to create opportunities for our workers to rise.

Secondly, and just as importantly, we need a cultural change, because fundamentally, this is about our values, about how we value people. And Singapore must always be a place where everyone can feel proud of what they do, where you are respected for your contributions and your character, and anyone can improve his life if he works hard, and everyone can hope for a better future.

Besides creating hope for the future, we must also give assurance to those in need, especially our seniors. By and large, seniors in Singapore are doing well. Many have savings and investments. Some are happily working, others are getting good support from their children. And we have good schemes to provide assurance in retirement. And the CPF scheme and home ownership are the twin pillars of our retirement adequacy.

Home ownership is critical. The Government has worked hard to help Singaporeans own their homes. Therefore, Singapore is not just a place to live, but a home. The CPF has also served us well and there are three good things about the CPF. Firstly, there is personal responsibility because with the CPF scheme, the more you work, the longer you work, the more you save and the more you'll have in retirement. Secondly, the CPF scheme is fair. Your savings are for your own retirement, not for someone else's. Thirdly, with CPF Life, this is for life, because CPF Life will pay you a stream of income as long as you live.

This is the CPF logo. You've all seen it, but if you look carefully, you'll see that it has got three keys inside it. And the three keys represent the three parties who have come together in order to make the CPF system and provide this social security for you.

Who are the three parties? One is yourself because when you work you earn a salary. From the salary, you pay your employee's CPF. The second key is your employer, because the employer pays into the CPF, into your CPF, the employer's CPF contribution over and above your wages because the Government required them by law to do it. So the third key is 'zheng hu' - the Government. The Government made this system, the Government set the rules. The Government also tops up directly into the CPF of many Singaporeans.

The CPF scheme is good, but it can be improved. It works well for most Singaporeans but not quite for all, especially the lower income. And also, it is not quite flexible enough. And I think we can and should improve the scheme further.

First of all, we should help the lower-income Singaporeans. With the CPF and HDB for the majority of the population, you can save enough for your retirement. But for a minority, 10, 20 percent, I think even if they are working, they may not accumulate enough CPF during their working lives. Some of them may not have bought an HDB flat, some may have no family support to fall back on and in their case, these individual efforts will not be enough. So, the Government and the society must help to do more, must do more to help them in their retirement. So, for this group, we should supplement their payouts from their own CPF savings with bonus payments from the Government.

The second thing we should do with the CPF scheme is to increase flexibility, to make it more flexible.

Now, my view is that the core purpose of the CPF should still be to provide a steady stream of income in old age, but I appreciate why some CPF members want to take more money out because they've been saving up over a lifetime of work, they want to use some of these savings. So after considering this for a long time and discussing it with my colleagues, I've decided I will change my view, I will adjust the policy. And I think we should allow people to the option to take out part of their CPF savings in a lump sum if they need to, but subject to some limits. The amount which you can take out cannot be excessive. For example, it can be up to 20 percent of the total that you have. And it should only be during retirement, 65 and beyond.

So the CPF and home ownership provide for our needs when we retire. They are good schemes, they work well for the majority of Singaporeans, but they are not one-size-fits-all policies. They offer different choices for people in different circumstances. But we are going to improve them further so that we can better support lower-income elderly who need more help. And we will make it more flexible so as to meet the needs of more Singaporeans, and give you greater assurance and more options in retirement.

We want Singapore to be the best place to live, work and play. We want this to be an outstanding city, well-planned, well-run, offering a high quality of life, full of buzz and vibrancy.

In the last 49 years, our physical transformation has been remarkable. Our Singaporean identity is strengthening, but keeping Singapore special is a journey without end. We'll work with Singaporeans to improve on what we have. We can do this and so much more to keep Singapore special, but what matters most is not what we build but the power of our human spirit, showing determination and resolve like our pioneers, aiming high and pushing ahead, as our young should, contributing in big ways and small to Singapore, no matter what our station in life.

We have all contributed to the Singapore Story. At the heart of the Singapore Story is our belief in Singapore, belief that we can turn vulnerability and despair into confidence and hope; belief that out of the trauma of separation, we could build a modern metropolis and a beautiful home; belief that whatever the challenges of this uncertain world we can thrive and prosper as one united people. Let this belief and spirit burn bright in each one of us and guide us forward for the next 50 years and more. Together, let us be the pioneers of our generation. Together, let us create a brighter future for all Singaporeans. Thank you and good night.

Appendix 14 Briefing note of Speech D in the post-test

Prime Minister Lee Hsien Loong (Singapore)

National Day Rally speech

At the ITE College Central, 2017

Background

National Day Rally (NDR) is an annual address that the Prime Minister of Singapore makes to the entire nation, on the first or second Sunday after National Day on 9 August. At the rally, the Prime addresses the nation on its key challenges and announce major policy changes.

Speech Summary

The speaker talks about building up preschools to ensure that every child starts well and has a bright future, fighting diabetes and making Singapore a Smart Nation by using IT to create opportunities for all Singaporeans.

Glossary

PUB: The Public Utilities Board is the Singaporean statutory board of the Ministry of the Environment and *Water* Resources responsible for ensuring a sustainable and efficient *water* supply

MAS: The *Monetary Authority of Singapore* is Singapore's central bank and financial regulatory authority.

Smart Nation: *Smart Nation* is the national effort of Singaporeans, businesses and government to support better living using technology, by having smarter ideas, apps and solutions.

Appendix 15 Script of speech D in the post-test

Good evening again. My fellow Singaporeans, we've had an eventful year. We've been busy in guarding against terrorism and strengthening our racial harmony making friends and cooperating with other countries, big and small.

Last year, I spoke about our economy: how we will be working to develop skills, to build capabilities, to promote entrepreneurship and take the economy to the next level. And I am happy to report good progress. We expect growth around 2.5 percent this year, higher than last year. Wages have been rising, gradually but steadily. And most encouragingly, productivity is improving. Last year, productivity went up by one percent, after several years of almost zero growth. And this year, we should do even better. And this is important because productivity is the key to our prosperity and to higher wages.

We still have work to do. I'd like to discuss three longer term issues that are important to the success and wellbeing of Singapore for the current generation and also for future generations.

One, building up our preschools so that every child, regardless of his family background, starts well and has a bright future. Two, fighting diabetes because many Singaporeans suffer from it, not only the old but increasingly younger people too. And three, making Singapore a Smart Nation, by using IT to the full, to create jobs and opportunities for all Singaporeans. Preschool, Diabetes and Smart Nation, these are the things we must do now, work on how to build our future so that Singaporeans can start right, stay healthy and live smart at every age.

Let me start with preschool. Here we are talking about infants to six year olds. Nowadays, even two month old babies are enrolled in an infant care. And that is part of preschool. Preschool is important to give our children a good start and the best chance to succeed in life.

We are doing three things to build up preschools and give young children a good start: More places for zero to four, better quality for five to six, raising the standard standing of teachers and carers. These changes will benefit all preschool kids. In addition, we are making a special effort for kids from low income and vulnerable families. These kids need more support, starting even earlier.

I've described what we are doing to develop our children in practical ways, but actually, we are emphasizing preschools to achieve a broader social purpose because access to affordable, quality preschools will help level the playing field for young children. Today, every child goes to a good school. We want every child to go to a good preschool so that all children, regardless of family background, have the best possible start in life.

My second topic tonight is our health, and specifically diabetes. You may not think diabetes is a major problem, but in fact, it is very serious in Singapore, particularly so for older people but increasingly for younger Singaporeans. Generally, Singaporeans think of ourselves as being fairly healthy. After all, we live quite long but it is not just about how many years you live, the quality of life matters greatly. And you know from your own experience with aging parents that when people grow old they suffer from many ailments and frailties, often dragging on for years.

What causes this ill health? One big reason is diabetes. And unfortunately, here compared to other developed countries, Singapore is almost world champion, just behind the US. Overall one in nine Singaporeans has diabetes. But the prevalence increases as you age. If you are my age, over 60, three in ten Singaporeans have diabetes.

The challenge with diabetes is that in the early stages, it is an invisible disease. You do not feel sick. There are hardly any symptoms. You may not even know that you have it. But if it is not treated, over time it can become very serious. If you look at the top causes of death in Singapore, diabetes does not appear there, but actually many common causes of death can be traced back to diabetes. So what can each one of us do?

Let me offer four suggestions. First, please get regular medical check-ups. Find out whether you have diabetes or ar are at risk. Do not take the attitude that it is better not to know. You must want to know, because if you know your condition, then you can do something about it.

My second suggestion, please exercise more. Exercise is good for you. It helps with your blood sugar, your blood pressure. It brings down your weight. It makes you feel better. But if you prefer something more fun, join a group activity.

My third piece of advice is to eat less and eat healthily. I was looking at some of my old school photos recently. And it reminded me that when I was in school, the children were not as tall, and some of us were quite scrawny. Perhaps we were not eating enough and were a little bit under-nourished.

My fourth piece of advice is to cut down on soft drinks. Soft drinks contain a lot of refined sugar, which is very bad for you. If you drink soft drinks every day, you are overloading your system with sugar and significantly increase your risk of diabetes.

We are scouting around for solutions. Some countries, several European ones, Mexico, Brunei, they've tried a sugar tax. UK and Chile also placed warning labels on drinks with high sugar content. But it is not clear yet if these measures work.

As a first step, we have got the soft drink producers to agree to reduce the sugar in all their soft drinks sold in Singapore. This will help. But ultimately, what to drink is a personal choice. The best is to drink plain water. Better still, drink PUB water.

I just described four simple things each one of us can do: get a check-up, exercise more, watch your diet, cut down on sugar. It requires commitment, adjustments to our habits, our lifestyles and diet. But the payoff is large and it can be done.

My third topic tonight is a Smart Nation. What is a Smart Nation about? Some think it is about each person owning two handphones, or having the fastest internet connection. Others talk about e-commerce, the Internet of Things, self-driving cars, artificial intelligence, big data. Those are all part of it, but not the whole story.

Smart Nation is about Singapore taking full advantage of IT. Using IT comprehensively to create new jobs, new business opportunities, to make our economy more productive, to make our lives more convenient, to make this an outstanding city in which to live, work and play.

We have a natural advantage. We are compact. We are highly connected. Our people are digitally literate. Our schools are teaching students basic computing and robotics. But while we have the right ingredients, we lag behind other cities in several areas. For example, electronic payments.

China has gone the furthest with e-payments. Indeed, in major Chinese cities, cash has become obsolete. Even debit and credit cards are becoming rare. Everyone is using WeChat Pay or AliPay and these apps are linked to your bank account.

In Singapore, we too have e-payments, but we have too many different schemes and systems that don’t talk to one another. We must simplify and integrate our systems. MAS has been working hard at this – integrating the different systems into one, so now, at last, we have one single unified terminal that can read different cards.

Another area where IT can help is public safety and security. Many cities already have comprehensive CCTV and sensor networks. And they also can integrate the inputs from all the sources, analyse and make sense of the information, and respond promptly if there is an incident or an emergency.

To do such Smart Nation projects, big or small, we need engineers, programmers, data analysts, technicians. We need people with the skills. We need managers with the understanding. We need leaders with the dare and the courage and the organizational ability to make it happen. When we started out with economic development, we put a lot of emphasis on engineering and science. In fact, when we gave scholarships, almost all the scholarships were for engineering. But in the last decade, two decades, the trend has shifted. It was balanced, from engineering to economics and liberal arts, which was a good sign.

Tonight, I have spoken about three things we are doing to build our future. Making our preschools better because we want every child to start well and have the best chance in life. Declaring war on diabetes so that we can stay well, live healthily and enjoy the fruits of our labors. Building a Smart Nation to create opportunities for all of us and keep Singapore a leading city in the world. Why are we so preoccupied with the future? Whom are we doing this for? Not just for ourselves, but for our children and future generations.

This is the Singapore of the last half century. Every generation striving and building for the next; keeping our eyes on tomorrow and investing in our children; undaunted by challenges and disruptions, instead, working together to overcome every obstacle, seize every opportunity and realize a bright future for all of us. Good night!

Appendix 16 Rating scales used in the current research

The rating scale for fluency of delivery

Score Range	Descriptors_Fluency of Delivery
90-100	Delivery is very fluent and disfluencies are rare.
80-89	Delivery is generally fluent, containing a small number of disfluencies.
70-79	Delivery is quite fluent and acceptable, with regular disfluencies.
60-69	Delivery is not fluent, containing frequent disfluencies which may impede comprehension.
59 and below	Delivery unacceptable, leading to the incomprehensibility of the message.

The rating scale for overall interpreting performance

Score Range	Descriptors_Overall Interpreting Performance
90-100	message delivered accurately with intended effect, delivery very fluent, target language use natural and idiomatic
80-89	the message generally delivered with the intended effect but with a few deviations that did not affect the overall meaning, delivery generally fluent with a few disfluencies, target language use overall natural and idiomatic with a few unnatural and incorrect usage
70-79	the message overall delivered but with regular deviations that affected the overall meaning, delivery is acceptable with regular disfluencies, target language use with regular unnatural and incorrect usage
60-69	message delivered inaccurately with frequent deviations that compromised overall message and coherence, delivery not fluent containing frequent disfluencies, target language use with frequent unnatural and incorrect usage
59 and below	message barely delivered and is consistent with the original, delivery inacceptable, target language use unnatural and incorrect

References

Ahrens, B. (2005). Prosodic phenomena in simultaneous interpreting: A conceptual approach and its practical application. *Interpreting*, *7*(1), 51–76.

Alexieva, B. (1988). Analysis of the simultaneous interpreter's output. In P. Nekeman (Ed.), *Translation, our future. Proceedings, XIth World Congress of FIT* (pp. 484–488). Maastricht: Euroterm.

Altman, J. (1994). Error analysis in the teaching of simultaneous interpretation: A pilot study. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (pp. 25–38). Amesterdam/Philadelphia: John Benjamins Publishing Company.

Anderson, J. R. (1983a). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Anderson, J. R. (1983b). *The Architecture of Cognition*. Cambridge: Harvard University Press.

Anderson, J. R. (2015). *Cognitive Psychology and its Implications* (8th ed.). New York: Worth Publishers.

Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap:Empirical research in simultaneous interpretation* (pp. 101–120). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Angelelli, C. V., & Jacobson, H. E. (2009). *Testing and assessment in translation and interpreting studies: A call for dialogue between research and practice*. Amsterdam/Philadelphia: John Benjamins Publishing.

Ankerstein, C. A. (2014). A psycholinguistic measurement of second language proficiency: The coefficient of variation. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 109–121). Bristol: Multilingual Matters.

Babcock, L., & Vallesi, A. (2017). Are simultaneous interpreters expert bilinguals, unique bilinguals, or both? *Bilingualism: Language and Cognition*, *20*(2), 403–417.

Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Bajo, M. T., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. G.-S. Salvador, & Y. Gambier (Eds.), *Translation in context* (pp. 127–142). Amsterdam/Philadelphia: Johan Benjamins Publishing Company.

Bakti, M. (2009). Speech disfluencies in simultaneous interpretation. In D. De Crom (Ed.), *Selected Papers of the CETRA Research Seminar in Translation Studies 2008* (pp. 1–18).

Becker, M., Schubert, T., Strobach, T., Gallinat, J., & Kühn, S. (2016). Simultaneous interpreters vs. professional multilingual controls: Group differences in cognitive control as well as brain structure and function. *NeuroImage*, *134*, 250–260.

Bendazzoli, C., Sandrelli, A., & Russo, M. (2011). Disfluencies in simultaneous interpreting: a corpus-based analysis. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 282–306). London/New York: Continuum.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn. Brain, Mind, Experience and School*. Washington D. C.: National Academy Press.

Brumfit, C. J. (1984). *Communicative methodology in language teaching : The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, *5*(4), 231–235.

Cai, R., Dong, Y., Zhao, N., & Lin, J. (2015). Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *Interpreter and Translator Trainer*, *9*(1), 104–120.

Caplan, D., Waters, G., & DeDe, G. (2007). Specialized verbal working memory for language comprehension. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 272–302). New York: Oxford University Press.

Cecot, M. (2001). Pauses in simultaneous interpretation: A contrastive analysis of professional interpreters' performances. *The Interpreters' Newsletter*, *11*, 63–85.

Cenoz, J. (1998). *Pauses and Communication Strategies in Second Language Speech*. Retrieved from https://eric.ed.gov/?id=ED426630

Chang, C. (2005). *Directionality in Chinese/English simultaneous interpreting: Impact on performance and strategy use (Doctoral dissertation)*. The University of Texas at Austin.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. *Psychological Review*, *92*, 195–222.

Chernov, G. V. (1979). Semantic aspects of psycholinguistic research in simultaneous interpretation. *Language and Speech*, *22*(3), 277–295.

Chernov, G. V. (2004). *Inference and Anticipation in Simultaneous Interpreting: A probability-prediction model*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cheung, A. K. F. (2007). The effectiveness of summary training in consecutive interpreting (CI) delivery. *Forum*, *5*(2), 1–23.

Cheung, A. K. F. (2013). Non-native accents and simultaneous interpreting quality perceptions. *Interpreting*, *15*(1), 25–47.

Chiaro, D., & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the world wide web. *Meta*, *49*, 278-293.

Chincotta, D., & Underwood, G. (1998). Simultaneous interpreters and the effect of concurrent articulation on immediate memory: A bilingual digit span study. *Interpreting*, *3*(1), 1–20.

Chmiel, A. (2018). In search of the working memory advantage in conference interpreting–Training, experience and task effects. *International Journal of Bilingualism*, *22*(3), 371–384.

Christoffels, I. K., De Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324–345.

Christoffels, I. K., De Groot, A. M. B., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism:Language and Cognition*, *6*(3), 201–211.

Clark, R. C. (2008). *Building expertise: Cognitive methods for training and performance improvement* (3rd ed.). New York: Pfeiffer.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information-processsing system. *Psychological Bulletin*, *104*(2), 163–191.

Cowan, N. (1997). *Attention and memory: An integrated framework*. New York and Oxford: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambiridge: Cambridge University Press.

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989–999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech.

Journal of the Acoustical Society of America, *111*(6), 2862–2873.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*(6), 445–464.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18.

Das, K. R., & Imon, A. H. M. R. (2016). A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, *5*(1), 5–12.

Davies, M., & Gardner, D. (2010). *A Frequency Dictionary of Contemporary American English:Word sketches,collocates and thematic lists*. London and New York: Routledge.

De Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, *13*(1), 420–442.

De Groot, A. M. B. (2000). A complex-skill approach to translation and interpreting. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting* (pp. 53–68). Amesterdam/Philadelphia: John Benjamins Publishing Company.

De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*(5), 600–629.

De Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning*, *47*(2), 215–264.

De Jong, N. (2013). How do I choose which measure of fluency to use? [Powerpoint slides]. Retrieved from https://prezi.com/qpi0fdpzxlqb/how-do-i-choose-which-measure-of-fluency-to-use/

De Jong, N. H., Schoonen, R., & Hulstijn, J. H. (2009). Fluency in L2 is related to fluency in L1. In *Seventh International Symposium on Bilingualism*. Utrecht, The Netherelands.

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.

Dechert, H. W., & Raupach, M. (1980). Towards a cross-linguistic assessment of speech production. Frankfurt: Peter Lang.

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.

Dong, Y. (2018). Complex dynamic systems in students of interpreting training.

Translation and Interpreting Studies, *13*(2), 185–207.

Dong, Y., & Cai, R. (2015). Working memory and interpreting: a commentary on theoretical models. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 63–81). Bristol, Buffalo and Toronto: Multlingual Matters.

Dong, Y., & Liu, Y. (2016). Classes in translating and interpreting produce differential gains in switching and updating. *Frontiers in Psychology*, *7*, 1297.

Dong, Y., Liu, Y., & Cai, R. (2018). How does consecutive interpreting training influence working memory: A longitudinal study of potential links between the two. *Frontiers in Psychology*, *9*, 1–12.

Dong, Y., & Xie, Z. (2014). Contributions of second language proficiency and interpreting experience to cognitive control differences among young adult bilinguals. *Journal of Cognitive Psychology*, *26*(5), 506–519.

Dong, Y., & Zhong, F. (2017). Interpreting experience enhances early attentional processing, conflict monitoring and interference suppression along the time course of processing. *Neuropsychologia*, *95*, 193–203.

Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, *25*(1), 11–28.

Ericsson, K. A. (2000). Expertise in interpreting: An expert-performance perspective. *Interpreting*, *5*(2), 187–220.

Ericsson, K. A. (2001). The path to expert golf performance: Insights from the masters on how to improve performance by deliberate practice. In *Optimising performance in golf* (pp. 1–57). Brisbane, Australia: Australian Academic Press.

Ericsson, K. A. (2002). Attaining excellence through deliberate practice: Insights from the study of expert performance. In *The pursuit ofexcellence in education* (pp. 21–55). Hillsdale: Erlbaum.

Ericsson, K. A. (2014). The acquisition of expert performance: An introduction to some of the issues. In K. Anders Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1–50). New York and London: Psychology Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Ericsson, K. A., & Smith, J. (Eds.). (1991). Prospects and limits of the empirical study of expertise: An introduction. In *Toward a general theory of expertise: Prospects and limits* (pp. 1–39). Cambridge: Cambridge University Press.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert adn exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*(1), 273–305.

Fabbro, F., & Darò, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain and Language*, (48), 309–319.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first-and second-language reading of fluent bilinguals. *Memory & Cognition*, *11*(6), 565–574.

Field, A. (2018). *Discovering statistics using IBM statistics* (5th ed.). London: Sage Publications.

Field, J. (2004). *Psycholinguistics: The key concepts*. London and New York: Routledge.

Fillmore, C. J. (2000). On fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 43–60). Michigan: The University of Michigan Press.

Foltz, P. W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 197–202.

Fortkamp, M. B. M. (1999). Working memory capacity and L2 speech production. *Communication and Cognition*, *32*, 259–296.

Fortkamp, M. B. M. (2000). *Working memory capacity and L2 speech production: An exploration study (doctoral dissertation)*. Federal University of Santa Catarina.

Fortkamp, M. B. M., & Bergsleithner, J. M. (2007). Relationship among individual differences in working memory capacity, noticing, and L2 speech production. *Signo*, *32*(52), 40–53.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, (18), 299–323.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent. In B. F. Freed (Ed.), *Second language acquisition in a study-abroad context* (pp. 123–148). Amsterdam: John Benjamins Publishing Company.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor: The University of Michigan Press.

Fu, R. (2012). Pausing in two-way E-C consecutive interpreitng-A contrastive analysis of trainee interpreters' performance (in Chinese). *Foreign Language Teaching and Research*, *44*(3), 437–447.

Fu, R. 符荣波. (2013). Directional effects on disfluencies in consecutive interpreting [口译方向性对译语非流利产出的影响]. *Modern Foreign Languages [现代外语]*, *36*(2), 198–205.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77–83.

García-Amaya, L. J. (2012). *Second language fluency and cognition: The study of Spanish second language development in an overseas immersion program and an* *at-home foreign language classroom (doctoral dissertation)*. Indiana University, Ann Arbor.

García, A. M. (2014). The interpreter advantage hypothesis: Preliminary data patterns and empirically motivated questions. *Translation and Interpreting Studies*, *9*(2), 219–238.

Gearhart, A., Booth, D. T., Sedivec, K., & Schauer, C. (2013). Use of Kendall's coefficient of concordance to assess agreement among observers of very high resolution imagery. *Geocarto International*, *28*(6), 517–526.

Gerver, D. (1969/2002). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 53–66). London and New York: Routledge.

Gile, D. (1991). A communication-oriented analysis of quality in nonliterary translation and interpretation. In M. L. Larson (Ed.), *Translation: Theory and practice, tension and interdependence* (pp. 188–200). Amsterdam/Philadelphia: John Benjamins Publishing.

Gile, D. (2001). L'évaluation de la qualité de l'interprétation en cours de formation. *Meta*, *46*(2), 379–393.

Gile, D. (2005). Teaching conference interpreting: A contribution. In M. Tennent (Ed.), *Training for the new millenium* (pp. 127–151). Amesterdam/Philadelphia: John Benjamins Publishing Company.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gile, D. (2011). Errors, omissions and infelicities in broadcast interpreting: Preliminary fndings from a case study. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies* (pp. 201–218). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*(3), 226–231.

Goldman-Eisler, F. (1968). *Psycholinguistics experiments in spontaneous speech*. London: Academic Press.

Gósy, M. (2007). Disfluencies and self-monitoring. *Govor*, *24*(2), 91–110.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202.

Grosjean, F. (1980). Temporal variables within and between languages. In H. W. Dechert & M. Raupach (Eds.), *Towards a cross-linguistic assessment of speech production* (pp. 39–53). Frankfurt: Lang.

Guará-Tavares, M. (2013). Working memory capacity and L2 speech performance in planned and spontaneous conditions: a correlational analysis. *Trabalhos Em Linguística Aplicada*, *52*, 9–29.

Guillot, M.-N. (1999). *Fluency and its teaching*. Clevedon, Philadelphia, Toronto and Sydney: Multilingual Matters.

Hamidi, M., & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta*, *52*(2), 276–289.

Han, C. (2015a). (Para)linguistic correlates of perceived fluency in English-to-Chinese simultaneous interpretation. *Journal of Comparative Literature and Translation*, *3*(4), 32–37.

Han, C. (2015b). Investigating rater severity/leniency in interpreter performance testing: A multifaceted Rasch measurement approach. *Interpreting*, *17*(2), 255–283.

Han, C. (2016). Investigating score dependability in English/Chinese interpreter certification performance testing: A generalizability theory approach. *Language Assessment Quarterly*, *13*(3), 186–201.

Han, C. (2018). Using rating scales to assess interpretation Practices , problems and prospects. *Interpreting*, *20*(1), 60–96.

Han, C., & Riazi, M. (2017). Investigating the effects of speech rate and accent on simultaneous interpretation: A mixed-methods approach. *Across Languages and Cultures*, *18*(2), 237–259.

Han, C., & Slatyer, H. (2016). Test validation in interpreter certification performance testing. *Interpreting*, *18*(2), 225–252.

Hargrove, P. M., & McGarr, N. S. (1994). *Prosody management of communication* *disorders*. Sandiego and California: Singular Publishing Group Inc.

Hatim, B., & Mason, I. (1997). *The translator as communicator*. London and New York: Routledge.

Hild, A. (2011). Effects of linguistic complexity on expert processing during simultaneous interpreting. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of porcess reserach: Integrative approaches in translation studies* (pp. 249–267). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hiltunen, S., Pääkkönen, R., Vik, G.-V., & Krause, C. M. (2016). On interpreters' working memory and executive control. *International Journal of Bilingualism*, *20*(3), 297–314.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Company.

Hu, W. 胡伟杰, & Wang, J. 王建勤. (2017). Predictive power of L2 oral cognitive fluency to L2 oral competence [第二语言口语认知流利性对口语能力的预测作用]. *Chinese Teaching in the World [世界汉语教学]*, *31*(1), 105–115.

Injoque-Ricle, I., Barreyro, J. P., Formoso, J., & Jaichenco, V. I. (2015). Expertise, working memory and articulatory suppression effect: Their relation with simultaneous interpreting performance. *Advances in Cognitive Psychology*, *11*(2), 56–63.

Jin, X. 金霞. (2011). L2 speaking span test and its predictive power of L2 oral fluency [二语口语广度测试及其对口语流利度的预测能力]. *Contemporary Foreign Languages Studies [当代外语研究]*, (1), 36–41.

Jin, X. 金霞. (2012). The effect of working memory capacity on written language production of second language learners [工作记忆容量限制对二语学习者口语产出的影响]. *Foreign Language Teaching and Research [外语教学与研究]*, *44*(4), 523–535.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs/New Jersey: Prentice-Hall Inc.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, & J. Towse (Eds.), *Variation in working memory* (pp. 21–48). Oxford: Oxford University Press.

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. *Routledge* (2nd ed.). New York and London: Routledge.

Kendall, M. G., & Babington Smith, B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, *10*(3), 275–287.

Kopczynski, A. (1994). Quality in conference interpreting: Some pragmatic problems. In S. Lambert & B. Moser-mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 87–99). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting*, *8*(1), 1–23.

Köpke, B., & Signorelli, T. M. (2012). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism*, *16*(2), 183–197.

Kormos, J. (2000). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition*, *22*(2), 145–167.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah and New Jersey: Lawrence Erlbaum Associates.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, (32), 145–164.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137–171.

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. In L. Gran & D. Snelling (Eds.), *The Interpreter's Newsletter* (pp. 13–21). Trieste: EUT.

Kurz, I. (1994). What do different user groups expect from a conference interpreter?

The Jerome Quarterly, *9*(2), 3–7.

Kurz, I. (1996). Special features of media interpreting as seen by interpreters and users. In *New horizons: Proceedings of the XIVth world congress of FIT* (Vol. 2, pp. 957–965). Melbourne: AUSIT.

Kurz, I. (2001). Conference interpreting: quality in the ears of the user. *Meta*, *46*(2), 394–409.

Kurz, I. (2003). Quality from the user perspective. In Á. C. Aís, M. F. Sánchez, & D. Gile (Eds.), *La evaluación de la calidad en interpretación: Investigación* (pp. 3–22). Granada: Editorial Comares.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852.

Lee, J. (2008). Rating scales for interpreting performance assessment. *Interpreter and Translator Trainer*, *2*(2), 165–184.

Leeson, R. (1975). *Fluency and Language teaching*. Clevedon, Philadelphia,Toronto and Sydney: Multilingual Matters.

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. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Michigan: The University of Michigan Press.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: MIT press.

Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, (22), 1–75.

Li, X. 李晓媛, & Yu, L. 俞理明. (2009). The role of key visuals in reducing working memory load during L2 listening comprehension [关键图示对二语听力理解中工作记忆负荷削减作用的研究]. *Modern Foreign Languages [现代外语]*, *32*(2), 149–157.

Lim, H., & Godfroid, A. (2014). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*(5), 1247–1282.

Lin, Y., Lv, Q., & Liang, J. (2018). Predicting Fluency With Language Proficiency, Working Memory, and Directionality in Simultaneous Interpreting. *Frontiers in*

Psychology, *9*(1543).

Liu, M. (2011). Methodology in interpreting studies: A methodological review of evidence-based research. In B. Nicodemus & L. Swabey (Eds.), *Advances in interpreting research: Inquiry in action* (pp. 85–119). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Liu, M. (2013). Design and analysis of Taiwan's interpretation certification examination. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting* (pp. 163–178). Frankfurt: Peter Lang.

Liu, M. (2015a). Assessment. In F. Pöchhacker (Ed.), *Routledge Encyclopedia of Interpreting Studies* (pp. 20–22). New York: Routledge.

Liu, M. (2015b). Certification. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 45–46). New York: Routledge.

Liu, M., & Chiu, Y. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting*, *11*(2), 244–266.

Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, *6*(1), 19–42.

Liu, M., Zhang, J., & Wu, S. (2008). Interpretation evaluation practices: Comparison of eleven schools in Taiwan, China, Britain, and the USA (in Chinese). *Compilation and Translation Review*, *1*(1), 1–42.

Macnamara, B. N., & Conway, A. R. A. (2016). Working memory capacity as a predictor of simultaneous language interpreting performance. *Journal of Applied Research in Memory and Cognition*, *5*(4), 434–444.

Marrone, S. (1993). Quality: A shared objective. *The Interpreters' Newsletter*, *5*, 35–41.

McLaughlin, B. (1995). Aptitude from an information-processing perspective. *Language Testing*, *12*, 370–387.

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter*, *10*, 89–102.

Mead, P. (2002). Exploring hesitation in consecutive interpreting: An empirical study. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st Century: Challenges and opportunities* (pp. 73–82). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Mead, P. (2005). Methodological issues in the study of interpreters' fluency. *The Interpreters' Newsletter*, *13*, 39–63.

Meisel, J. M. (1987). A note on second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic Models of Production*. Norwood, NJ: Ablex Publishing Corporation.

Mellinger, C. D., & Hanson, T. A. (2019). Meta-analyses of simultaneous interpreting and working memory. *Interpreting*, 21(2), 165–195.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, London, and New Delhi: Sage Publications.

Meuleman, C., & Van Besien, F. (2009). Coping with extreme speech conditions in simultaneous interpreting. *Interpreting*, *11*(1), 20–34.

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology*, *130*(2), 163–168.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640.

Miyake, A., & Shah, P. (1999). Models of working memory: Mechanisms of active maintenance and executive control. Cambridge and New York: Cambridge University Press.

Mizera, G. J. (2006). *Working memory and L2 oral fluency (doctoral dissertation)*. University of Pittsburgh, Pennsylvania, USA.

Mizuno, A. (2017). Simultaneous interpreting and cognitive constraints. *Bulletin of College of Literature*, *58*, 1–28.

Monsell, S. (1996). Control of mental processes. In V. Bruce (ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove: Erlbaum.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342.

Morales, J., Padilla, F., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Simultaneous interpretation selectively influences working memory and attentional networks. *Acta Psychologica*, *155*, 82-91.

Morris, N., & Jones, D. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*, 111–121.

Moser-Mercer, B. (1996). Quality in interpreting: Some methodological issues. *The Interpreters' Newsletter*, *7*, 43–55.

Moser-Mercer, B. (2008). Skill Acquisition in Interpreting. *The Interpreter and Translator Trainer*, *2*(1), 1–28.

Moser-Mercer, B., Frauenfelder, U. H., Casado, B., & Künzli, A. (2000). Searching to define expertise in interpreting. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives* (pp. 107–131). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Moser, P. (1996). Expectations of users of conference interpretation. *Interpreting*, *1*(2), 145–178.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos*, (24), 69–104.

Nation, P. (1989). Improving speaking fluency. *System*, *17*(3), 377–384.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581.

Padilla, F., Bajo, M. T., & Macizo, P. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tommola (Ed.), *Topics in Interpreting Research* (pp. 61–72). Turku: University of Turku.

Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, (10), 319–335.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amesterdam and Philadelphia: John Benjamins Publishing Company.

Petite, C. (2005). Evidence of repair mechanisms in simultaneous interpreting: A corpus-based analysis. *Interpreting*, *7*(1), 27–49.

Pinget, A.-F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365.

Pio, S. (2003). The relation between ST delivery rate and quality in simultaneous interpretation. *The Interpreters' Newsletter*, *12*, 69–100.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting. *Translation and Interpreting Studies*, *11*(2), 202–224.

Plonsky, L. (2015). *Advancing quantitative methods in second language research*. New York and London: Routledge.

Pöchhacker, F. (1994). Quality assureance in simultaneous interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims and visions* (pp. 233–242). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pöchhacker, F. (1995). Slips and shifts in simultaneus interpretation. In J. Tommola (Ed.), *Topics in Interpreting Research* (pp. 73–90). Turku: University of Turku.

Pöchhacker, F. (1997). "Clinton speaks German": A case study of live broadcast simultaneous interpreting. In M. Snell-Hornby, Z. Jettmarová, & K. Kaindl (Eds.), *Translation as intercultural communication: Selected papers from the EST Congress, Prague 1995* (pp. 207–216). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pöchhacker, F. (2001). Quality assessment in conference and community interpreting. *Meta*, *46*(2), 410–425.

Pöchhacker, F. (2004). *Introducing Interpreting Studies*. London and New York: Routledge.

Pöchhacker, F. (2012). Interpreting quality: Global professional standards. In W. Ren (Ed.), *Interpreting in the age of gobalization: Proceedings of the 8th National*

Conference and International Forum on Interpreting [全球化时代的口译：第八届全国口译大会暨国际研讨会论文集] (pp. 305–318). Beijing: Foreign Language Teaching and Research Press [外语教学与研究出版社].

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*(77), 97–131.

Pradas Macías, M. (2006). Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency. *Interpreting*, *8*(1), 25–43.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 263–270). Berlin: De Gruyter Mouton.

Raupach, M. (1987). Procedural learning in advanced learners of a foreign language. In J. A. Coleman & R. Towell (Eds.), *The advanced language learner*. London: CILT.

Rehbein, J. (1987). On fluency in second language speech. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 97–106). Norwood, NJ: Ablex Publishing Corporation.

Rennert, S. (2010). The impact of fluency on the subjective assessment of interpreting quality. *Interpreters Newsletter,* (15), 101–115.

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.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology*, *124*(2), 207–231.

Rosiers, A., Woumans, E., Duyck, W., & Eyckmans, J. (2019). Investigating the presumed cognitive advantage of aspiring interpreters. *Interpreting*, *21*(1), 115–134.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*(3), 395–412.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 45–66). New Jersey: Ablex.

Sawyer, D. B. (2004). *Fundamental aspects of interpreter education: Curriculum and assessment*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*(4), 357–385.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 750–755). Mahwah: Lawrence Erlbaum Associates.

Schneider, E., & Zuccoloto, A. (2007). E-prime 2.0 [computer software]. Pittsburg, PA:

Psychological Software Tools.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories-new models. *Interpreting*, *13*(2), 176–204.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: The University of Michigan Press.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York and London: Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics*, *54*(2), 79–95.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*(2), 173–199.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, *33*(4), 644–653.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385.

Segalowitz, S. J., Segalowitz, N., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*(1), 53–67.

Seleskovitch, D. (1986). Comment: Who should assess an interpreter's performance. *Multilingua*, *5*(4), 236.

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*(3), 383–408.

Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis*. Amsterdam/Philadelphia: John Benjamins Publishing.

Setton, R., & Dawrant, A. (2016). *Conference interpreting: A trainer's guide*. Amesterdam/Philadelphia: John Benjamins Publishing Company.

Shadrick, S. B., & Lussier, J. W. (2009). Training complex cognitive skills: A theme-based approach to the development of battlefield skills. In K. A. Ericsson (Ed.), *The development of professional expertise: Toward measurement of expert prformance and design of optimal learning environment* (pp. 286–311). Cambiridge: Cambridge University Press.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology*, *125*(1), 4–27.

Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreters' Newsletter*, *12*, 37–49.

Shreve, G. M., Lacruz, I., & Angelone, E. (2011). Sight translation and speech disfluency: Performance analysis as a window to cognitive translation process. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 93–120). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Signorelli, T. M., Haarmann, H. J., & Obler, L. K. (2011). Working memory in simultaneous interpreters: Effects of task and age. *International Journal of Bilingualism*, *16*(2), 198–212.

Simone, R. (1995). *Fondamenti di Linguistica*. Bari: Editori Laterza.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14.

Snellings, P., van Gelderen, A., & de Glopper, K. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning*, *52*(4), 723–754.

Stainton, R. J. (2006). Preface. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. xiii–xiv). Oxford: Blackwell Publishing.

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, *30*(1), 79–99.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Talmy, L. (2008). Aspects of attention in language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 37–48). New York and London: Routledge.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Timarová, Š., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting*, *16*(2), 139–168.

Timarová, Š., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2015). Simultaneous interpreting and working memory capacity. In A. Ferreira & J. W. Schwieter (Eds.), *Psycholinguistic and cognitive inquiries into translationand interpreting* (pp. 101–126). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters' Newsletter*, (10), 103–127.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics*, *40*(2), 117–150.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–119.

Tzou, Y.-Z., Eslami, Z. R., Chen, H.-C., & Vaid, J. (2011). Effect of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: Evidence from Mandarin–English speakers. *International Journal of Bilingualism*, *16*(2), 213–227.

Wang, B., & Li, T. (2015). An empirical study of pauses in Chinese-English simultaneous interpreting. *Perspectives*, *23*(1), 124–142.

Wang, J. (2013). Bilingual working memory capacity of professional Auslan/English interpreters. *Interpreting*, *15*(2), 139–167.

Wang, J. (2016). The relationship between working memory capacity and simultaneous interpreting performance: A mixed methods study on professional Auslan/English interpreters. *Interpreting*, *18*(1), 1–33.

Wang, J., & Napier, J. (2015). Measuring bilingual working memory capacity of professional Auslan/English interpreters: A comparison of two scoring methods. *The Interpreters' Newsletter*, *19*, 45–62.

Weissheimer, J. (2007). *Working memory capacity and the development of L2 speech production: An exploratory study (doctoral dissertation)*. Federal University of Santa Catarina.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics*, *22*(1), 1–22.

Wen, H., & Dong, Y. (2019). How does interpreting experience enhance working memory and short-term memory: A meta-analysis. *Journal of Cognitive Psychology, 31(8), 769-784.*

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs." *Psychological Research*, *63*(3), 212–233.

Xiao, R., Rayson, P., & McEnery, T. (2009). *A Frequency Dictionary of Mandarin Chinese: Core vocabulary for learners*. London and New York: Routledge.

Xu, H. 徐海铭. (2010). Pauses in conference consecutive interpreting from English into Chinese: An empirical study [汉英交替传译活动中的口译停顿现象实证研究——以国际会议职业口译受训译员为例]. *Foreign Languages Research [外语研究]*, *119*(1), 64–71.

Yu, W., & van Heuven, V. J. (2017). Predicting judged fluency of consecutive interpreting from acoustic measures: Potential for automatic assessment and pedagogic implications. *Interpreting*, *19*(1), 47–68.

Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, *2*(309), 1–9.

Zeng, J. 曾记, & Hong, M. 洪媚. (2012). A study on self-repairs of trainee interpreters in Chinese-English consecutive interpreting [学生译员汉译英连传口译中的自我修正研究]. *Foreign Languages and Their Teaching [外语与外语教学]*, *264*(3), 68–72.

Zhang, W. Z. 张文忠. (1999a). Research into second language oral fluency: A survey [国外第二语言口语流利性研究现状]. *Foreign Language Teaching and*

Research [外语教学与研究], *2*, 41–48.

Zhang, W. Z. 张文忠. (1999b). The theoretical model for the development of second language oral fluency [第二语言口语流利性发展的理论模式]. *Modern Foreign Languages [现代外语]*, *84*(2), 202–217.

Zhang, W. Z. 张文忠, & Wu, X. 吴旭东. (2001). Development of L2 oral fluency: A quantitative analysis [第二语言口语流利性发展定量研究]. *Modern Foreign Languages [现代外语]*, *24*(4), 341–351.