

**Investigating Trainee Interpreters' Cognitive Load in Simultaneous  
Interpreting and Sight Translation**

**By**

**Duoduo Li**

**Doctor of Philosophy Languages, Cultures and Religions**

**September 2023**

**Faculty of Literature and Languages**

**University of Stirling**

## Acknowledgements

Embarking on this PhD journey just as the global pandemic began presented unique challenges, yet also fostered my personal growth. Beginning my academic pursuit in October 2019, I was aware of the impending challenges. However, the rapid spread of COVID-19 unexpectedly magnified these difficulties, especially with the unforeseen trial of being separated from my parents for nearly four years. Despite the many adversities, the past four years have been incredibly rewarding. Originating from a background in Translation and Interpreting from my undergraduate and master's studies, the interdisciplinary nature of my PhD project introduced me to the intricacies of Cognitive Psychology. This not only involved mastering domain-specific knowledge but also necessitated that I start from scratch with statistics, learning many data analysis software, and acquainting myself with various experimental equipment. I felt like a sponge, constantly soaking in vast arrays of previously unfamiliar knowledge. I am elated by my perseverance that has brought me to this point today.

First, I would like to extend my profound gratitude to Professor Raquel de Pedro Rico, who stepped in as my main supervisor during a pivotal moment in my PhD journey. In my second year, I changed my main supervisor. I am deeply appreciative of her decision to join our supervisory team at such a critical time. When I transitioned to her supervision, the extensive revisions required for my previously written content felt quite overwhelming. However, with numerous discussions and her guidance, I felt a steady improvement in my work. Her attention to detail in my study and her timely feedback have been pivotal in refining my writing. Every time I express my gratitude, she humbly responds, "that is why I am here." I truly respect her expertise in the fields of interpretation and translation.

Next, I must express immense gratitude to my second supervisor, Dr Jan Kuipers. Without his support, my PhD journey would have been tenfold more challenging. He patiently explained complex statistical concepts to me, acquainted me with the nuances of experimental equipment, and mentored me in experimental design and the use of analytical tools like R. Our consistent weekly meetings during the research design phase helped me transition from a beginner to being nearly proficient with both the software and equipment. When I expressed my appreciation for the considerable time he dedicated to supporting me, Jan always responded that, considering my initial unfamiliarity with the subject, he believed it was essential to offer

more support to reduce my struggles. Throughout the journey, Jan has always motivated me, pointing out positive strides in my writing and the noticeable progress from where I began. I am genuinely grateful for all his guidance and encouragement.

My heartfelt appreciation goes to all the students who participated in my experiments and the peers who aided in making these experiments feasible. During the pandemic, the approval of lab experiment was repeatedly postponed, and there were concerns about a lack of participants. Fortunately, your assistance ensured the smooth completion of essential components of my research.

I would like to extend my gratitude to Dr Fiona Barclay, who was the previous PGR Director in my department. She provided invaluable support during the challenging period of my supervisor transition. Her persistent care for my well-being has continued even now, despite no longer overseeing our doctoral program. She still maintains contact with me and offers advice on my academic progress.

Thanks to my close friend, Lani Nelson, our friendship started during our master course. Over the years, she has accompanied me through many important occasions. In my PhD journey, she has played a supportive role, offering valuable insights. Our friendship is something I truly cherish.

I also want to thank my colleagues, Mrs Caroline Marley and Professor Jim Mills, at the University of Strathclyde. Over the past four years, I have been an academic tutor for a programme there, and balancing this role with my studies was not easy. Their understanding and acknowledgement have been invaluable.

Lastly, but most importantly, to my parents – the cornerstone of all my endeavours. Their unwavering love and support have been my guiding light. My father has always supported every decision I have made, continually offering encouragement. My mother has been my emotional anchor in my life. She unconditionally believes in me, consistently offering encouragement. She always stands by me with understanding, especially in my toughest emotional times. My mother has also consistently shared wisdom and thoughts that has shaped my perspective on life. To me, she is the strongest and tenderest person in the world. Mum and Dad, I love you immensely.

As my PhD journey nears its conclusion, the dawn of a new chapter in my life emerges. This doctoral experience has fortified my professional expertise and emotional resilience, equipping me with the confidence to face forthcoming challenges. Once again, I would like to express my gratitude to everyone who has assisted me along this journey.

## **Abstract**

While SI and ST have distinct mechanisms, they both exert significant cognitive demands on interpreters. Each mode has its own unique input method, and this difference underscores the importance of a comparative analysis. This study investigates the effects of delivery rate and text difficulty on trainee interpreters' cognitive load in simultaneous interpreting (SI) and sight translation (ST), comprising both an online experiment and a lab experiment. To delve deeply into this, a mixed-method approach was adopted. Data collection encompassed both quantitative and qualitative methods, including self-assessments, pupillometry, galvanic skin response (GSR), holistic accuracy assessments, and retrospective interviews.

Online study findings indicated a higher cognitive load for trainees in SI than in ST. In SI, both delivery rate and text difficulty influenced the trainees' perceived cognitive load and their accuracy scores, with delivery rate showing a stronger impact. In contrast, for ST, while both factors affected perceived cognitive load and accuracy, text difficulty had a more pronounced effect on trainees' output accuracy. Lab study confirmed greater cognitive load in SI compared to ST, with delivery rate significantly influencing pupil size in SI. Both delivery rate and text difficulty influenced accuracy in SI, reaffirming the stronger impact of delivery rate. In ST, although no significant changes in pupil size were observed due to either factor, accuracy was influenced by both, with text difficulty showing a more pronounced effect. Furthermore, the performance-based measurement of cognitive load, represented by trainees' accuracy scores, consistently confirmed these findings, whereas physiological measurements, like pupil size, showed mixed results. No significant differences were observed in GSR across various conditions in both SI and ST tasks, and no correlation was identified between GSR and either pupil size or accuracy scores. This research offers empirical evidence for the cognitive comparison between sight translation and simultaneous interpreting, providing some valuable insights for pedagogical practices.

## Table of Contents

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Acknowledgements.....                                                      | 2  |
| Abstract.....                                                              | 5  |
| List of Tables and Figures.....                                            | 10 |
| List of Abbreviations .....                                                | 12 |
| Chapter 1 Introduction .....                                               | 13 |
| 1.1 Research Background .....                                              | 13 |
| 1.2 Research Aims .....                                                    | 16 |
| 1.3 Research Questions .....                                               | 17 |
| 1.4 Independent, Dependent and Covariate Variables .....                   | 18 |
| 1.5 Structure of the Thesis .....                                          | 18 |
| Chapter 2 Literature Review .....                                          | 21 |
| 2.1 Simultaneous Interpreting and Sight Translation .....                  | 21 |
| 2.1.1 The Concept of Simultaneous Interpreting and Sight Translation.....  | 21 |
| 2.1.2 The Comparative Studies of SI and ST in Interpreting Studies.....    | 25 |
| 2.2 Cognitive Load in SI and ST Studies.....                               | 30 |
| 2.2.1 Mental Load vs. Cognitive Load .....                                 | 30 |
| 2.2.2 Cognitive Load Concept in Interpreting and Translation Studies ..... | 32 |
| 2.2.3 Previous Literature on Cognitive Load in SI and ST.....              | 33 |
| 2.3 Methods of Measuring Cognitive Load .....                              | 44 |
| 2.3.1 Analytical Method .....                                              | 44 |
| 2.3.1.1 A Brief Overview of the Analytical Method .....                    | 44 |
| 2.3.1.2 The Tightrope Hypothesis and Effort Model .....                    | 45 |
| 2.3.2 Subjective Method .....                                              | 50 |
| 2.3.2.1 A Brief Overview of the Subjective Method.....                     | 50 |
| 2.3.2.2 Task Load Index.....                                               | 50 |
| 2.3.3 Performance Method.....                                              | 51 |
| 2.3.4 Psycho-Physiological Method .....                                    | 52 |
| 2.3.4.1 A Brief Overview of the Psycho-physiological Method.....           | 52 |
| 2.3.4.2 The Application of Pupillometry.....                               | 52 |
| 2.3.4.3 The Application of GSR.....                                        | 55 |
| 2.4 Experimental Variables.....                                            | 57 |
| 2.4.1 A Brief Description of the Input Variables.....                      | 57 |
| 2.4.2 Delivery Rate as An Input Variable.....                              | 58 |
| 2.4.3 Text difficulty as An Input Variable .....                           | 59 |
| 2.4.4 Working Memory Capacity as A Covariate Variables .....               | 64 |

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Chapter 3 Methodology .....                                                      | 66 |
| 3.1 Mixed-methods Approach for Assessing Cognitive Load .....                    | 66 |
| 3.2 A Within-subject Design.....                                                 | 66 |
| 3.3 Statistical Method for Analysing Experimental Results .....                  | 67 |
| 3.4 Methodology for Experiment One .....                                         | 69 |
| 3.4.1 Subjective Method: The Self-Assessment of Perceived Cognitive Load .....   | 70 |
| 3.4.2 Performance Method: Accuracy Evaluation .....                              | 70 |
| 3.4.3 Working Memory Capacity Measurement: Corsi Span Test .....                 | 74 |
| 3.5 Methodology for Experiment Two .....                                         | 75 |
| 3.5.1 Performance Method: Accuracy Evaluation .....                              | 76 |
| 3.5.2 Psycho-physiological Method: Pupillary Measurement .....                   | 77 |
| 3.5.3 Psycho-physiological Method: GSR Measurement .....                         | 78 |
| 3.5.4 Combining GSR and Pupil Dilation for Psycho-physiological Measurement..... | 78 |
| 3.5.6 Retrospective Method: Interview .....                                      | 79 |
| 3.5.7 Working Memory Capacity Measurement: Word Span Test .....                  | 80 |
| Chapter 4 Experiment One.....                                                    | 83 |
| 4.1 Research Questions .....                                                     | 83 |
| 4.2 Independent, Dependent and Covariate Variables .....                         | 83 |
| 4.3 Hypothesis in Experiment One .....                                           | 83 |
| 4.4 Research Design.....                                                         | 84 |
| 4.4.1 Experiment Instruments .....                                               | 84 |
| 4.4.1.1 Online Experiment Platform .....                                         | 84 |
| 4.4.1.2 Self-assessment of Perceived Cognitive Load .....                        | 84 |
| 4.4.1.3 Corsi Span Test .....                                                    | 85 |
| 4.4.2 Participants.....                                                          | 85 |
| 4.4.3 Experimental Stimuli .....                                                 | 86 |
| 4.4.3.1 Brief on Source Texts.....                                               | 86 |
| 4.4.3.2 Evaluation of Source Text Difficulty .....                               | 87 |
| 4.4.3.3 Readability Measurement .....                                            | 87 |
| 4.4.3.4 Word Frequency Measurement .....                                         | 88 |
| 4.4.3.5 Constraints on Delivery Rate and Presentation of Source Texts .....      | 89 |
| 4.4.3.6 Speaker in the SI Task.....                                              | 90 |
| 4.4.3.7 Stimuli for Corsi Task.....                                              | 90 |
| 4.4.4 Experiment Procedures .....                                                | 90 |
| 4.4.5 Data Treatment.....                                                        | 92 |
| 4.4.5.1 SI and ST Accuracy Scoring .....                                         | 92 |
| 4.4.5.2 Self-assessment Data .....                                               | 93 |

|                                                                            |     |
|----------------------------------------------------------------------------|-----|
| 4.4.5.3 Corsi Task Scoring .....                                           | 93  |
| 4.5 Result of Experiment One.....                                          | 93  |
| 4.5.1 Performance Analysis: SI and ST Accuracy Results .....               | 93  |
| 4.5.2 Subjective Analysis: Perceived Cognitive Load Results .....          | 95  |
| 4.5.3 The Relationship between Accuracy and Perceived Cognitive Load ..... | 97  |
| 4.5.4 Working Memory Capacity Results.....                                 | 98  |
| 4.6 Discussion .....                                                       | 98  |
| 4.7 Limitations and Implications .....                                     | 102 |
| Chapter 5 Experiment Two .....                                             | 104 |
| 5.1 Research Questions, Research Variables and Hypothesis .....            | 104 |
| 5.1.1 Research Questions .....                                             | 104 |
| 5.1.2 Research Variables.....                                              | 104 |
| 5.1.3 Hypothesis in Experiment Two .....                                   | 105 |
| 5.2 Research Design.....                                                   | 105 |
| 5.2.1 Experiment Instruments .....                                         | 105 |
| 5.2.1.1 Eye tracker .....                                                  | 105 |
| 5.2.1.2 Eye-tracker Calibration Settings .....                             | 106 |
| 5.2.1.3 E-Prime for Tobii .....                                            | 107 |
| 5.2.1.4 GSR Device .....                                                   | 107 |
| 5.2.2 Participants.....                                                    | 108 |
| 5.2.3 Experimental Stimuli .....                                           | 109 |
| 5.2.3.1 Brief on Source Texts.....                                         | 109 |
| 5.2.3.2 Evaluation of Source Text Difficulty .....                         | 110 |
| 5.2.4 Constraints on Delivery Rate and Presentation of Source Texts .....  | 116 |
| 5.2.5 Speaker in the SI Task .....                                         | 116 |
| 5.2.6 Stimuli for Word Span Test .....                                     | 116 |
| 5.2.7 Fixed Questions in Semi-structured Interviews.....                   | 117 |
| 5.3 Experiment Procedures .....                                            | 117 |
| 5.4 Data Treatment.....                                                    | 118 |
| 5.4.1 Pupil Size Data and GSR Data.....                                    | 118 |
| 5.4.2 SI and ST Accuracy Scoring.....                                      | 118 |
| 5.4.3 Word Span Task Scoring .....                                         | 119 |
| 5.5 Results of Experiment Two .....                                        | 119 |
| 5.5.1 The Pupil Size Results .....                                         | 119 |
| 5.5.2 ST and SI Accuracy Results .....                                     | 122 |
| 5.5.3 The GSR Results.....                                                 | 125 |
| 5.5.4 The Working Memory Capacity Results .....                            | 125 |

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| 5.6 Discussion .....                                                    | 126 |
| 5.6.1 The Discussion of Research Questions .....                        | 126 |
| 5.6.2 The Discussion of Working Memory Capacity Results .....           | 135 |
| 5.6.3 The Application of Galvanic Skin Response in Lab Experiment ..... | 137 |
| 5.6.4 Further Discussion of the Interview .....                         | 138 |
| Chapter 6 Conclusion.....                                               | 140 |
| 6.1 Summary of the Research Findings .....                              | 140 |
| 6.2 Pedagogical Implications on SI and ST Training .....                | 141 |
| 6.3 Significances and Limitations .....                                 | 143 |
| 6.4 Future Avenues of the Research .....                                | 145 |
| References.....                                                         | 147 |
| Appendices.....                                                         | 166 |
| Appendix 1 Online Study (Experiment One) Consent Form .....             | 166 |
| Appendix 2 Lab Study (Experiment Two) Consent Form .....                | 167 |
| Appendix 3 Online Experiment Tutorial (in Chinese).....                 | 168 |
| Appendix 4 NASA Task Load Index (TLX) .....                             | 172 |
| Appendix 5 Marking Criteria for Each Score Range .....                  | 173 |
| Appendix 6 Experiment One Results.....                                  | 174 |
| Appendix 7 Experiment Two Results .....                                 | 178 |

## List of Tables and Figures

|                                                                                            |     |
|--------------------------------------------------------------------------------------------|-----|
| Table 1 Comparative Studies related to SI and ST .....                                     | 29  |
| Table 2 Research on Cognitive Load in Sight Translation and Simultaneous Interpreting..... | 34  |
| Table 3 Text Difficulty Measurements in Translation and Interpreting Studies .....         | 62  |
| Table 4 Profile of Participants .....                                                      | 86  |
| Table 5 Condition Pair Comparison in ST.....                                               | 95  |
| Table 6 Condition Pair Comparison in SI.....                                               | 95  |
| Table 7 Condition Pair Comparison in ST.....                                               | 97  |
| Table 8 Condition Pair Comparison in SI.....                                               | 97  |
| Table 9 Correlation Result between Accuracy Score and Self-assessment Score.....           | 97  |
| Table 10 Correlation Result of Working Capacity .....                                      | 98  |
| Table 11 Profile of Participants .....                                                     | 109 |
| Table 12 Five Bilinguals' Profile.....                                                     | 114 |
| Table 13 Bilingual Rater's Evaluation of Difficulty of Source Texts .....                  | 115 |
| Table 14 ANOVA: Post Hoc Comparisons – Text Type.....                                      | 115 |
| Table 15 Fixed Effects Parameter Estimates in SI Task.....                                 | 124 |
| Table 16 Fixed Effects Parameter Estimates in ST Task.....                                 | 125 |
| Table 17 Correlation Result of GSR Data .....                                              | 125 |
| Table 18 Correlation Result of Working Memory Capacity .....                               | 126 |
|                                                                                            |     |
| Figure 1 Readability Test Accuracy Scores.....                                             | 87  |
| Figure 2 Readability Test Accuracy Scores.....                                             | 88  |
| Figure 3 Word Frequency Score .....                                                        | 89  |
| Figure 4 Frowny Face Means the Answer is Incorrect .....                                   | 90  |
| Figure 5 Mean Score of ST and SI Accuracy .....                                            | 94  |
| Figure 6 Mean Perceived Cognitive Load Scores.....                                         | 96  |
| Figure 7 The Tobii 1750 Eye-tracker.....                                                   | 106 |
| Figure 8 Correct Eye Positioning in the Track Status Box.....                              | 106 |
| Figure 9 Calibration Results .....                                                         | 107 |
| Figure 10 “Correct setting of GSR” (Biopac Student Lab, 2014, p. 6) .....                  | 108 |
| Figure 11 Readability Test Accuracy Scores.....                                            | 111 |

|                                                               |     |
|---------------------------------------------------------------|-----|
| Figure 12 Readability Test Accuracy Scores .....              | 111 |
| Figure 13 Word frequency scores .....                         | 112 |
| Figure 14 Word Frequency Evaluation of Normal Text I.....     | 113 |
| Figure 15 Word Frequency Evaluation of Difficult Text H ..... | 113 |
| Figure 16 Mean Pupil Size Value in SI .....                   | 120 |
| Figure 17 Mean Pupil Size Value in ST .....                   | 120 |
| Figure 18 Mean Pupil Size Value in ST and SI .....            | 121 |
| Figure 19 Mean Accuracy Score in SI.....                      | 122 |
| Figure 20 Mean Accuracy Score in ST .....                     | 123 |
| Figure 21 Mean Accuracy Score in SI and ST .....              | 123 |

## List of Abbreviations

|          |                                                                                        |
|----------|----------------------------------------------------------------------------------------|
| SI       | Simultaneous Interpreting                                                              |
| ST       | Sight Translation                                                                      |
| SIT      | Simultaneous Interpreting with Text                                                    |
| GSR      | Galvanic Skin Response                                                                 |
| TAPs     | Think-Aloud Protocols                                                                  |
| EEG      | Electroencephalogram                                                                   |
| fMRI     | Functional Magnetic Resonance Imaging                                                  |
| fNIRS    | Functional Near-Infrared Spectroscopy                                                  |
| WM       | Working Memory                                                                         |
| WMC      | Working Memory Capacity                                                                |
| CLT      | Cognitive Load Theory                                                                  |
| C-E/ E-C | Chinese-English/English-Chinese                                                        |
| BC       | Before Common Era                                                                      |
| AOIs     | Area of Interest                                                                       |
| EOIs     | Errors, Omissions and/or Infelicities                                                  |
| NASA-TLX | National Aeronautics and Space Administration Task Load Index                          |
| PGR      | Psychogalvanic Reflex                                                                  |
| EDR      | Electrodermal Response                                                                 |
| AIIC     | International Association of Conference Interpreters                                   |
| NICTA    | Australia's Information and Communications Technology Research<br>Centre of Excellence |
| ILO      | International Labour Organization                                                      |
| ANOVA    | Analysis of Variance                                                                   |
| NN       | Normal Difficulty Text, Normal Delivery                                                |
| NF       | Normal Difficulty Text, Fast Delivery Rate                                             |
| DN       | Difficult Text, Normal Delivery Rate                                                   |
| DF       | Difficult Text, Fast Delivery Rate                                                     |

## **Chapter 1 Introduction**

This study investigates the effects of delivery rate and text difficulty on trainee interpreters' cognitive load in simultaneous interpreting (SI) and sight translation (ST), comprising both an online experiment and a lab experiment. The decision to initiate with the online experiment was largely influenced by the need to navigate the constraints presented by the pandemic. Additionally, it offered valuable insights and experience that aided the subsequent lab experiment. Recognising the limitations of online experiments, the lab experiments presented a more refined and comprehensive experimental measure. While the overarching aim remains consistent, the experimental methods employed in the two experiments have some differences, and the specifics will be elaborated upon in the methodology chapter. The core focus of this study lies in examining the effects of delivery rate and text difficulty variations on the cognitive load of trainee interpreters in simultaneous interpreting and sight translation tasks. To delve deeply into this, a mixed-method approach was adopted. Data collection tools cover both quantitative and qualitative domains, including self-assessments, pupillometry, galvanic skin response (GSR), holistic accuracy assessments, and retrospective interviews. Such an investigation could potentially provide pedagogically valuable insights, leading to improved translation and interpreting training methods and better outcomes. Furthermore, owing to its interdisciplinary nature, this study synergizes Interpreting and Translation Studies with Cognitive Psychology Studies. Such a blend holds potential to advance academic discourse on interpreting and translation, while also offering novel perspectives on bilingual task-oriented cognitive research.

This introduction provides an overview of the research. It begins with the research background, setting the context and highlighting its relevance as well as the potential contributions. Next, the research objectives and questions are presented, defining the direction of this study. The research design section then outlines the chosen methodologies and their rationale. Lastly, the structure of the dissertation is detailed, offering a guide to the following chapters.

### **1.1 Research Background**

The period since the 1970s has been particularly significant for the development of translation and interpreting. This era saw a rapid maturation of Translation and Interpreting Study as a

formalised discipline. Scholars, noticing the details and complexities, began to incorporate foundational theories and methodologies from an array of disciplines to enhance the depth and breadth of their understanding. Over the subsequent decades, Translation and Interpreting Study has burgeoned into a vibrant hub for multidisciplinary research. Exemplifying the dynamic, interdisciplinary inclination of modern research, contemporary investigation into the intricate processes of interpreting now melds theories from diverse areas such as cognitive science, neuroscience, and even computer science. These theories, in conjunction with advanced technological tools, have the potential to foster fresh insights and delineate emerging trajectories in the evolving landscape of interpreting research.

Considering the advancements in interdisciplinary studies, it has become increasingly clear among scholars that translating and interpreting are not mere linguistic activities. These are, in essence, sophisticated cognitive endeavours where the brain grapples with multiple tasks in real-time, as observed in translation and interpreting tasks. By the end of the 20th century, there was a noticeable shift towards a cognitive approach in Translation and Interpreting Studies (Korpala 2016). Cognitive research in translation and interpreting has drawn from relevant theories in cognitive science. This discipline investigates the complicated mechanisms of information representation, processing, and transformation through faculties such as perception, language, memory, and reasoning within our biological nervous system (Li 2019). It has provided insights into how the human brain processes multilingual information, especially when multitasking (Albir and Alves 2008). Simultaneous interpreting and sight translation tasks, which combines comprehension and production tasks, presents significant cognitive challenges. Furthermore, the cognitive complexities heighten significantly when the interpreting task bridges languages with starkly different features (Seebert 2011). A prime example of this is the challenge presented when translating between English and Chinese, two languages that differ greatly in their syntactic, semantic, phonetic, and morphological constructs.

As a form of interpretation, simultaneous interpreting was defined by the European Commission (2018) as “a mode of interpreting in which the speaker makes a speech, and the interpreter reformulates the speech into a language his audience understands at the same time (or simultaneously)”. To put it simply, simultaneous interpreting is the process of listening, comprehending, and immediately interpreting what a speaker says into another language. On the other hand, “sight translation is a form of transposing a written text in the source language

into the target language orally, and its sub-processes include reading and comprehending the source language text as well as reformulating and producing it in the target language” (Li, 2019, p.13). Sight translation, as described, is very similar to a normal reading-out-loud mode (Shreve et al. 2010). The primary similarity between simultaneous interpreting and sight translation is their production of oral output. Conversely, the key difference between these two modes is the nature of their input: simultaneous interpreting relies on aural input, whereas sight translation is based on visual input. Both simultaneous interpreting and sight translation are widely used in various situations.

Though simultaneous interpreting and sight translation differ in their mechanisms, they both impose profound cognitive demands on practitioners. Sight translation, as a variant of simultaneous interpreting (Carpenter and McDonald 1981; Lambert 2004; Song 2010), remains an essential pedagogical tool in simultaneous interpreting training; this is because it helps trainees hone crucial interpreting skills such as deverbalization and syntactic linearity (Seleskovitch 1978; Mikkelsen 1993; Viaggio 1995; Agrifoglio 2004; Song 2010). A comparative study of simultaneous interpreting and sight translation could potentially provide valuable insights into the role of sight translation within simultaneous interpreting training. In addition, the cognitive load of an interpreter has been a topic of interest for an extended period in cognitive studies. When processing demand exceeds available capacity, the interpreter may miss information in the source language, which may be detrimental to the interpretation content. Such “source language overflow” is very common because in this multi-task system, as “the portion of an interpreter’s limited cognitive capacity devoted to performing an interpreting task” (Chen, 2017, p. 640). Such challenges are unique, as few tasks match the cognitive demands of interpreting, wherein “no physical activity is involved or need be accomplished, no instruments can be of help, everything goes on in the mind” (Riccardi et al., 1998, p. 97). Plevioets and Defrancq (2018) highlighted that research on cognitive load in interpreting often centres on instances where processing demands exceed capacity, affecting interpretation quality. Understanding cognitive load plays an important role in interpreting research. On the one hand, it offers invaluable insights into the myriad challenges that interpreters cope with daily. The level of cognitive load directly reflects the difficulty of the task, which is useful for selecting tasks that meet test requirements in interpretation education (Liu and Chiu 2009) and further aid development of training programmes. On the other hand, this exploration of translation and interpreting through the perspective of cognitive science not only deepens our understanding but also bridges the fields of interpreting and cognitive science.

In its continuous evolution, translation studies, while drawing from foundational theories of other fields, has also significantly expanded its use of advanced technological tools for data collection. Recent advancements in both technology and cognitive science research methodologies provide unprecedented opportunities to understand the complexities of cognitive load in interpreters and translators. Tools such as think-aloud protocols (TAPs), key-logging, electroencephalogram (EEG), galvanic skin response (GSR), functional magnetic resonance imaging (fMRI) have found application in empirical studies, significantly enhancing our capability to observe and measure cognitive processes during translation. While eye-tracking, which utilizes specific devices to capture individuals' eye movements, has brought forth innovative research methodologies in translation studies (Albir and Alves 2008), pupil dilation, also referred to as variations in pupil size, has emerged as a particularly valuable metric. Rooted in the eye-mind hypothesis, it is posited that a direct correlation exists between one's visual fixation and cognitive engagement (Jakobsen 2017). Thus, by observing changes in pupil size, we can glean insights into the cognitive activities of interpreters. Over the past decade, pupil dilation has become increasingly prominent in translation studies, especially as the field has welcomed more physiological measurements into its research purview. Another physiological measure is the GSR. This metric devices the fluctuations in the skin's electrical conductivity, which intrinsically varies with arousal levels. Given the potential of GSR to reflect changes in cognitive load, it has been extensively employed in diverse academic and commercial research endeavours. In this study, pupillary measurement and GSR measurement, were employed to observe the variations in cognitive load of trainee interpreters, providing a richer set of physiological data for data collection.

## **1.2 Research Aims**

Recognizing the distinct challenges posed by each mode, simultaneous interpreting (SI) and sight translation (ST) differ in their primary input methods. While SI is characterized by auditory input, ST is primarily based on visual input. However, both methods result in an oral output. Researchers have suggested that cognitive processes may vary between reading comprehension and listening comprehension (Patterson and Shewell 1987; Shallice 1987; Michael et al. 2001). This difference in input modes makes a comparison between the two especially worthy of study. Despite ST serving as a bridge between translation and interpreting,

and despite many interpreting training modules using ST as a pedagogical instrument for SI training, it remains under-explored, especially in studies comparing it to SI. Given this context, it becomes valuable to investigate the difference of trainee interpreters' cognitive load between SI and ST task.

From a Cognitive Psychology perspective, task difficulty is believed to directly influence an individual's cognitive load, while time pressure might have an indirect effect (Monod and Kapitaniak 1999; Galy et al. 2012). In Interpreting and Translation Studies, both text difficulty and delivery rate are seen as significant input variables potentially impacting an interpreter's performance (Gile 1995; Setton 1999; Pöchhacker 2016). Furthermore, as a core concept of cognitive load research, working memory capacity has been suggested by several researchers (e.g., Darò 1994; Chincotta and Underwood 2006; Christoffels et al. 2006; Macnamara and Conway 2016) to impacts the output of interpreters. Based on this rationale, we incorporated working memory capacity as a covariate variable. This approach aids in determining whether variations in trainees' interpreting performances can be attributed to differences in their working memory capacity. However, despite these theories, there remains a dearth of empirical evidence on the topic (Gile 2021). Thus, this study aims to gain an understanding of how text difficulty and delivery rate, treated as independent input variables, along with working memory capacity as a covariate variable, influence the trainee interpreters' cognitive load during ST and SI tasks.

### **1.3 Research Questions**

#### **Research Questions for Experiment One:**

**Question 1:** Do text difficulty and time pressure influence trainees' cognitive load in SI tasks?

**Question 2:** Do text difficulty and delivery rate influence trainees' cognitive load in ST tasks?

**Question 3:** Do trainees' cognitive load differ between ST and SI tasks?

**Question 4:** Does working memory capacity, when considered as a covariate variable, correlate with the cognitive load (accuracy result) of interpreting trainees in SI and ST tasks?

#### **Research Questions for Experiment Two:**

**Question 1:** How does text difficulty and delivery rate impact on trainees' cognitive load during an SI task?

**Question 2:** How does text difficulty and delivery rate impact on trainees' cognitive load during a ST task?

**Question 3:** How does cognitive load differ between ST and SI tasks among interpreting trainees?

**Question 4:** Does the impact of text difficulty and time pressure on cognitive load differ between ST and SI tasks performed by interpreting trainees?

**Question 5:** Does working memory capacity, influence the cognitive load (accuracy) of interpreting trainees in SI and ST tasks?

#### **1.4 Independent, Dependent and Covariate Variables**

- Independent Variables:
  - a. Task with the levels: sight translation (ST) and simultaneous interpreting (SI)
  - b. Delivery rate with the levels: normal vs. fast
  - c. Text difficulty with the levels: normal vs. difficult
- Dependent Variables:
  - a. ST and SI performance accuracy score (in both Experiment One and Two)
  - b. Perceived cognitive load self-assessment score (in Experiment One)
  - c. Pupil size (in Experiment Two)
  - d. GSR (in Experiment Two)
- Covariate Variables: Working Memory Capacity (WMC)

#### **1.5 Structure of the Thesis**

This thesis is constituted of six chapters: introduction (Chapter One), literature review (Chapter Two), methodology (Chapter Three), Online Experiment (Chapter Four), Lab Experiment (Chapter Five), General Discussion and Conclusion (Chapter Six).

Chapter One introduces the foundational aspects of the research, presenting the background that motivates this study. The focus then shifts to the research objectives and questions, elucidating the broader aims of the investigation. This chapter also touches upon the potential contributions of this research to the existing body of knowledge. Furthermore, a glimpse into

the research design and methodology is provided, culminating in an overview of the structural layout of this thesis.

Chapter Two presents a literature review that delves deep into the realms of SI and ST, providing definitions and drawing comparisons based on existing literature. It further elaborates on the concept of cognitive load in interpreting studies, covering its definition, associated models, and the methodologies employed for measurement. Within this context, the chapter sheds light on input variables, with a particular focus on the influence of delivery rate and text difficulty. Moreover, the role of working memory capacity (WMC) is elucidated, positioning it as a covariate variable in the exploration and understanding of the subject at hand.

Chapter Three provides an in-depth account of the methodological choices underpinning the research. Emphasis is placed on the constructs utilized for measuring cognitive load within the domain of interpreting studies. The methodology incorporates an array of tools and techniques: pupillometry, GSR, self-assessments and the assessment of SI and ST performance accuracy. The chapter culminates with a segment on the adoption of retrospective interviews, which serves as an additional lens to comprehend participants' experiences and perspectives.

Chapter Four delineates the design and execution of the online experiment (Experiment One). After presenting an in-depth look at the research design, it transitions to describe participant profiles, experimental stimuli, and data treatment. The results section illuminates findings on accuracy, WMC and perceived cognitive load outcomes. This leads into a discussion that interweaves the results with existing literature, concluding with an overview of the online experiment's limitations.

Chapter Five transitions to the lab experiment (Experiment Two), recognizing the limitations of the prior experiment. Methodological adjustments are made to enhance the robustness and validity of the study. It begins with an articulation of the research design, proceeding to detail participant backgrounds and experimental stimuli. Subsequent sections present the results, illuminating insights on accuracy, cognitive load, and other variables. The discussion segment not only examines the implications of the findings, especially concerning delivery rate, text difficulty, and related parameters, but also intertwines these with relevant literature for a holistic understanding.

Chapter Six firstly revisits the research questions and methods, then integrates the findings from both experiments, providing a comprehensive perspective. Emphasis is placed on the study's implications and contributions to the broader scholarly community, which includes summarizing the pedagogical implications of SI and ST training. The chapter concludes by showing the research significance, limitations, and suggesting potential directions for future studies.

## **Chapter 2 Literature Review**

Translating a source text into a target language, whether through written translation or verbal interpretation, has been a longstanding human practice. Scholars have been studying simultaneous interpreting and sight translation since the last century, but the study of the cognitive process of these two modes has only begun in recent decades. This chapter offers a thorough review of existing literature pertinent to this research project, primarily presenting the current state of knowledge in the field. This literature review delves into several key concepts, including simultaneous interpreting, sight translation, cognitive load and input variables. Through this review, the intent is to identify gaps in the current body of knowledge, underscore the importance of the research aims, and lay the groundwork for the formulation of research methods and theoretical frameworks.

Initially, the chapter delves into the definitions of SI and ST, supplemented by comparative studies of these two modes. This is followed by a detailed exploration of the context and definition of cognitive load, which includes an overview of the four methods used to measure it. The relationship between cognitive load and both SI and ST is examined, along with a review of previous studies that connect SI, ST, and cognitive load. Subsequently, the chapter presents a theoretical discourse on input variables. Given this study's particular emphasis on the effects of variations in speech rate, text difficulty in the source text, and working memory capacity, there is also a comprehensive review of past research exploring the impact of these variables on the cognitive load during SI and ST studies.

### **2.1 Simultaneous Interpreting and Sight Translation**

This section provides a comprehensive overview of SI and ST by presenting their definitions, exploring the backgrounds of both concepts, highlighting their similarities and differences, and reviewing relevant literature on comparative studies of the two modes.

#### **2.1.1 The Concept of Simultaneous Interpreting and Sight Translation**

Oral-based interpretation/translation is an ancient human practice. It is regarded as a special form of Translation, which predates the invention of writing and written translation (Pöchhacker 2001). The terminology used in Germanic, Scandinavian, and Slavic languages to

refer to an individual engaged in oral-based interpretation can be linked back to Akkadian, an ancient Semitic language from Assyria and Babylonia, dating back to around 1900 Before Common Era (BC). Pöchhacker (2009, p. 11) further defined Interpreting as “a form of Translation which a first and final rendition in another language is produced on the basis of a one-time presentation of an utterance in a source language”.

The history of simultaneous interpreting can be traced back to the early 20th century, with the first recorded use of the technique taking place at the International Labour Organization (ILO) in Geneva in 1927 (Baigorri 2014). However, the development of modern simultaneous interpreting as we know it today was largely influenced by the events of World War II and the subsequent rise of international organizations such as the United Nations. During the war, simultaneous interpreting was used by military personnel to facilitate communication between allied nations. The technique was also used at the Nuremberg Trials in 1945, where interpreters used headsets and microphones to provide real-time interpretation for the international audience (Gaiba 1998; Matasov 2017). After the war, simultaneous interpreting became more widely used in international conferences and meetings, with the first United Nations General Assembly in 1946 featuring the use of simultaneous interpretation in multiple languages (Albl-Mikasa and Tiselius 2021). The demand for simultaneous interpreting continued to grow, leading to the development of specialized equipment such as soundproof booths and advanced microphone systems.

As previously noted, simultaneous interpreting entails the real-time interpretation of speech into another language, enabling the audience to comprehend the message concurrently with the speaker’s delivery. Simultaneous interpreting is a very complex multi-tasking activity, it requires more cognitive demanding than consecutive interpreting or sight translation, because interpreters are speaking at the same time as listening and analysing (Gile 1995). Simultaneous interpreting requires a high level of linguistic and cognitive skills, as well as specialized equipment such as soundproof booths, headsets, and microphones. Interpreters must be able to listen to and process the speaker’s words quickly and accurately, while also producing an equivalent message in the target language at the same time. Hence, interpreters must possess robust language processing skills, either to swiftly dissect language segments or to employ a ‘sentence-driven’ approach, ensuring complete and accurate interpretation of the input they receive. The advantage of SI is it happens in real-time. Thus, SI is widely used in international conferences because it can meet the needs of multiple languages and allow delegates to choose

their favourite language to follow the proceedings. As globalization has deepened and more and more international conferences have been held, the demand for simultaneous interpreters has also increased. Many universities and institutions have opened an SI training course.

On the other hand, sight translation consists of converting written text from the source language to the target language orally. With the distinct combination of visual input and oral output, ST is often viewed as a blend between translation and interpreting (Dragsted and Hansen 2009). “It is generally assumed that oral interpretation and written translation are two fundamentally different processes, and strategies adopted by interpreters are also different from those adopted by translators” (Li, 2019, p.14). Sight translation appears to be “closer to interpreting than to translation because the oral modality carries an expectation of speedy delivery” (Dragsted and Hansen, 2009, p. 590). Lambert (2004, p. 298) also defines ST as “a variant of oral interpretation”. Although the source text continues to be visually accessible to the sight translators, it requires simultaneously comprehending the source language input and producing the output, which is similar to the work practice of SI (Gile, 1995, p. 169-170). As Lambert (2004, p. 298) suggested, “ST seems to have more in common with SI, given the number of variables involved – time stress, anticipation, reading for idea closure, not to mention the oral nature of the task – factors that are either absent in written translation, or present only to a limited degree”. While there is no consensus among scholars, Cenkova (2015, p. 375) advocated for the term “sight interpreting” over sight translation due to the latter’s lack of precision. Considering the demands of real-time processing, the practice aligns more closely with the characteristics of interpreting than translation, making “sight interpreting” a more accurate label for reflecting the nature of this specific interpreting mode (Cenkova 2015). Although ST leans more towards interpretation, it cannot be entirely detached from the domain of written translation nor solely categorized as a subfield of interpreting studies. As Baker (1998) argued, scholars should avoid creating divisions between the disciplines of translation and interpreting; instead, the goal should be to promote unity between the two, working towards their collective maturation. To holistically study sight translation, it is imperative to approach it from both the perspectives of written translation and interpreting. In doing so, especially within the context of a comparative study between SI and ST, it underscores the meaningfulness and richness of such an analysis, highlighting the unique intricacies of each mode while emphasizing their shared attributes.

In daily life, sight translation is a technique used by professional sight translators in a variety of settings, including press conference, community-based institutions, legal and medical contexts, where written documents need to be quickly and accurately translated for an audience. From the perspective of interpreting and translation training, sight translation has been employed in capability assessments to evaluate if candidates can rapidly comprehend the core of a text and convey its significance (Russo 2011). In addition, ST often seen as an effective precursor to SI training (Agrifoglio 2004; Mikkelsen 1993; Viaggio 1995; Song 2010). Practicing sight interpreting can enhance trainees' speed, prepare them for simultaneous interpreting tasks, and improve their ability to swiftly identify core information in the source text (Cenková 2015). ST also suggested as a foundational pedagogical tool in Simultaneous Interpreting with Text<sup>1</sup> (SIT) training. Sight translation also is a cognitively demanding task (Agrifoglio 2004). When it comes to cognitive processing, interpreters working with visual input, as opposed to auditory, benefit from the ability to dictate their own pace. However, the continuous presence of the text in sight translation can intensify cognitive demands, as the ever-visible text heightens the potential for lexical and syntactic interference (Gile 2009).

In summary, when discussing the modes of simultaneous interpreting and sight translation, several similarities can be identified:

- Oral Output: Both SI and ST require the interpreter to process the acquired information and subsequently convey it orally.
- Real-time delivery: On some occasions, both modes require the interpreter to deliver the message in real-time. In SI, the interpreter delivers the translation while the speaker is still speaking. In ST, the interpreter reads the written text and delivers the translation orally in real-time.
- Predictive Skills: In both SI and ST, interpreters must anticipate the upcoming content from the speaker or the text due to the real-time nature of these modes. This ability to foresee aids in ensuring smooth and continuous delivery.

On the other hand, the differences between simultaneous interpreting and sight translation can be identified are as follows:

---

<sup>1</sup> Simultaneous interpreting with text (SIT) is a mode that source language is presented both aurally and visually to the interpreters during the task.

- **Input Modality:** SI involves aural input as interpreters listen to the speaker's speech, while ST demands visual input with interpreters reading the source text.
- **Unit of Processing:** In ST, interpreters can view text at least at the sentence level, providing a more stable context. Conversely, in SI, they deal with spoken words that come and go quickly, without the stability of a fixed text.
- **Sources of Interference:** Both modes are subject to environmental influences. However, in SI, the interference may arise from the speaker's accent or issues with interpreting equipment. In contrast, during ST, visual challenges related to the layout or presentation of the source texts are more prevalent.

### **2.1.2 The Comparative Studies of SI and ST in Interpreting Studies**

While simultaneous interpreting and sight translation exhibit notable differences in their approaches and demands, the academic landscape is relatively sparse in terms of empirical studies that delve into a direct comparison between the two. This gap underscores the importance of further research to truly understand the nuances and distinct characteristics of each mode. A brief summary of the comparative studies between SI and ST is provided in Table 1 at the end of section 2.1.2.

Viezzi (1989) compared the information retention after listening, reading, simultaneous interpreting and sight translation. According to his research, retention rates are lower after sight translation compared to simultaneous interpreting. This suggests that the difference between these two modes of translation is not solely based on the way information is perceived or comprehended. Viezzi borrowed the theory from Graik and Lockhart (1972) that retention of information is influenced by the duration and depth of cognitive processing to further explain this finding. Interpreters may use different strategies in SI and ST because of the distinct input approach between these two modes. These different processing strategies impact retention rates. Furthermore, this research also drew the conclusion that retention is lower when there are more morphosyntactic transformations needed to translate the source language to the target language. Both sight translation and simultaneous interpreting have a cost in terms of information retention, which is related to the degree of morphosyntactic differences between the two languages. Greater retention occurs when the two languages involved in the translation are structurally similar. While Viezzi's study primarily focused on the subjects' retention of information following both SI and ST, his findings were beneficial and contributed to the field.

Firstly, this study utilized information retention as the parameter, which is an appropriate representation of the cognitive processes involved in the tasks being studied. Differences in the levels of information retention between sight translation and simultaneous interpreting suggest that the cognitive processes involved in each are distinct. The various processing processes required in each type of modes are dissimilar and may lead to varying degrees of reduction in cognitive capacity. Secondly, information retention is not only affected by the translation process itself, but also by the structural differences between the two languages. The morpho-syntactic differences between the Chinese and English languages are considerable. These disparities represent a crucial factor that triggers complications in the translation process, whether in sight translation or simultaneous interpreting, the difficulty of the source text's syntax can have a substantial impact on the translation outcome.

Another paper that involves the comparison of simultaneous interpreting and sight translation performance is an article published by Lambert in 2004. This research examined the processes involved when participants engage in sight translation compared to sight interpretation and simultaneous interpretation, with a specific focus on whether the visual presentation of the material being interpreted improves or hinders performance. The findings suggest that both sight translation and sight interpretation significantly outperform simultaneous interpretation, implying that visual access to the content being interpreted does not hinder a person's ability to listen and speak concurrently. Instead, it may even enhance their performance. According to Lambert (2004), the advantage of sight translation over simultaneous interpreting lies in the reduced processing difficulty, resulting in less interference in ST. However, this study does not detail the evaluation criteria for SI and ST performance as well as the detailed scoring method. Another factor that may affect the reliability of Lambert's findings is the possible fatigue experienced by the interpreters. This is because all participants in the study performed the tasks in the same order without any counterbalancing, which could have introduced additional variables affecting the results.

It is noteworthy to mention the article published by Agrifoglio (2004), which contains a comparison between sight translation and simultaneous interpreting performance. That study assessed the performance of six professional interpreters across those three modes, drawing the conclusions that each method has its unique difficulties and requires distinct levels of effort, and the interpreters' cognitive demands in the sight translation are no less significant than those involved in simultaneous or consecutive interpreting. Particularly, in this research, participants'

performance was analysed “based on failures (errors) observed in each mode” (Agrifoglio, 2004, p. 51), and errors were categorized into meaning errors and expression errors. Meaning errors, including changes in meaning and incomplete sentences, mostly resulted in information loss, except when redundant omissions were strategically used. Expression errors encompassed syntax, style, terminology, collocations, calques, and grammatical issues. The findings also show that the types of errors observed in sight translation were significantly different from those in simultaneous and consecutive interpreting. All interpreters exhibited fewer meaning problems but more expression issues in sight translation compared to the other two modes. Agrifoglio (2004) further argued that memory overload accounted for most errors in simultaneous interpreting, while the constant presence of the source text in sight translation caused difficulties, impacting target language expression and the balance between silent reading and oral translation. Visual distraction appears to have a stronger impact than audio interference, and even though the source text is always available, sight translators still need to depend on short-term memory, particularly when the grammatical structures between the two languages differ significantly. Although the data pool is quite limited, this paper, for the first time in comparative empirical research on simultaneous interpreting and sight translation, categorizes types of performance errors and provides some suggestions about the differences in cognitive demands caused by different types of input.

Chmiel et al. (2020) explored the interference of the source language (measured on the lexical and structural level) on forty-seven advanced interpreting trainees in both sight translation (visual modality) and simultaneous interpreting (auditory modality). The lexical analysis involved cognates and interlingual homographs within the sentences, while the structural analysis focused on sentences with agentless passive structures. The research incorporated two more dependent variables, namely time lag (operationalised as ear-voice span in SI and eye-voice span in ST) and translation time. The study’s results showed no task effect on the use of cognates, as they were used in a similar way in both ST and SI. Moreover, no specific preference was noted for cognate or non-cognate solutions in either task. An overwhelming majority of passive constructions were maintained by the participants (over 80% in both tasks), indicating that the trainees may lack the experience to make stylistically elegant and natural translation choices. Instead, they tend to choose the simplest solution due to limited resources, using the source language structure in the target language as it is grammatically correct. This strategy seems driven by the desire to save effort. Interference-based solutions occurred more frequently in ST than in SI, but only when it led to incorrect translations (homographs). With

respect to the temporal measures, the expected longer time lag in sight translation compared to simultaneous interpreting was confirmed. This could be due to the fact that SI is not paced by the interpreter, causing trainees to begin interpreting sooner in SI than in ST, perhaps to prevent overloading their memory buffer. Moreover, translation time was longer in SI than in ST, possibly due to the increased cognitive load required in SI. Chmiel et al. (2020) also pointed out that given the high cognitive demands of both ST and SI, future research should further investigate this aspect. It is noteworthy that this article incorporated a comparison of strategies into the performance comparison between sight translation and simultaneous interpreting under varying circumstances. The results also substantiate the existence of differing cognitive models and strategic approaches between the two. The researchers use translation time as a dependent variable lends a unique perspective to this research, which provide a new angle for further comparing the SI and ST cognitive processes. For instance, the amount of time pressure for performing the tasks (i.e., when time pressure is treated as an independent variable), how SI and ST would perform differently is worthy of further investigation.

Drawing comparisons from aforementioned studies is challenging due to the diverse factors involved. However, it is clear that the cognitive demands and strategies utilised in simultaneous interpreting and sight translation differ, with each's unique input modes introducing distinct challenges. A research gap exists in product-oriented comparative studies between sight translation and simultaneous interpreting, especially those focusing on the English-Chinese language pair, which are notably scarce. Additionally, the range of independent variables considered in studies of both modes is limited, encompassing aspects such as time pressure and text difficulty. Table 1 presents a general review of the comparative studies between SI and ST performance.

**Table 1 Comparative Studies related to SI and ST**

| Scholar                 | Language                               | Comparison Mode                  | Sample                                                 | Warming-up                                  | Independent variables                                                 | Dependent variables                           |
|-------------------------|----------------------------------------|----------------------------------|--------------------------------------------------------|---------------------------------------------|-----------------------------------------------------------------------|-----------------------------------------------|
| Viezzi (1989)           | English - Italian;<br>French - Italian | Listening,<br>Reading,<br>SI, ST | 12 trainee students;<br>9 professional<br>interpreters | No                                          | No                                                                    | Information<br>retention                      |
| Lambert (2004)          | French - English                       | SI, SIT, ST                      | 14 trainee students                                    | ST:5 mins<br>SIT: 10 mins<br>SIM: no        | No                                                                    | Performance<br>rates                          |
| Agrifoglio (2004)       | English - Spanish                      | SI, CI, ST                       | 6 professional<br>interpreters                         | ST: less than 5<br>mins<br>SI: no<br>CI: no | No                                                                    | Meaning<br>errors;<br>expression<br>errors    |
| Chmiel et al.<br>(2020) | English - Polish                       | SI, ST                           | 47 trainee students                                    | No                                          | Interlingual<br>homograph;<br>cognate; agentless<br>passive structure | Performance;<br>time lag;<br>translation time |

## **2.2 Cognitive Load in SI and ST Studies**

As mentioned before, both SI and ST are tasks with high cognitive demands. In this section, we will first delve into the general concept of cognitive load, discussing its essence and the rationale behind adopting this label. From there, we will narrow our focus to explore how the concept of cognitive load is particularly understood and applied in Translation and Interpreting studies. Subsequently, we will discuss the approaches employed in measuring cognitive load within this research domain. Lastly, we will conduct a review of past SI and ST studies that have directly investigated cognitive load, revealing trends and gaps in this crucial area of research.

### **2.2.1 Mental Load vs. Cognitive Load**

The terms ‘mental load’ and ‘cognitive load’ have been employed variously and often interchangeably, with no strict definition (Chen 2017). However, what distinguishes these labels? Which one aligns more closely with our research focus? To answer these questions, we must first delve into the definitions of both load-related terms.

The term ‘mental load’ refers to the cumulative effort required for managing the remembering of things (Ryder and Slater 2018). Analysing and measuring mental load has become a crucial topic in contemporary basic and applied psychological sciences. Examples of mental load include scenarios when you strive to remember something but struggle, or when task demands exceed your mental capacity. These experiences demonstrate the mental expenditure involved in task performance, a concept that forms the essence of mental load research (Bertelson et al. 1992). In applied sciences, professionals assess an individual’s ability to handle task demands, thereby estimating the mental load associated with the task. These evaluations influence design considerations, training, and personnel selection. Bertelson et al. (1992, p. 267) suggested that the analysis of mental load “can hence be viewed as an estimate of the cost of mental operations, and explication of the constraints imposed by these costs on the ability of a performer to cope with the demands of a task”. This definition of mental load, predominantly used in applied psychology can be linked to the concept of cognitive load. As elaborated by Paas et al. (2003, p. 64), “Mental load is the aspect of cognitive load that originates from the interaction between task and subject characteristics..... mental load can be determined on the basis of our current knowledge about task and subject characteristics”. Building on this understanding, it is noted

that mental load “provides an indication of the expected cognitive capacity demands and can be considered an a priori estimate of the cognitive load” (Paas et al., 2003, p. 64). Thus, mental load forms part of cognitive load measurement, representing the degree of mental burden individuals experience when completing a task. Essentially, before measuring how learners utilize their working memory or cognitive resources, we first need to understand the level of their mental load.

The idea of cognitive load originated in the 1950s alongside the development of working memory models. It is predicated on the notion that human cognitive resources are finite and thus can be overwhelmed by excessive demands (Welford 1952; Miller 1956; Broadbent 1958). Building on this foundational understanding, Cognitive Load Theory (CLT) was developed by Sweller in 1988. It was established upon the widely accepted model of human information processing, which is published by Richard Atkinson and Richard Shiffrin in 1968. According to the CLT, cognitive load can be identified in three different forms: intrinsic load, extrinsic load and germane load.

Intrinsic Load refers to the load induced by the intrinsic nature of the items being processed (Galy et al., 2012, p. 270), it is the load resulting from the inherent complexity of the learning task (Chandler and Sweller 1996; Sweller 1998) and directly affected by task difficulty (Bucks and Seljos 1994). This type of load depends on two factors, namely the element interaction of the task and the learner’s prior knowledge (Moreno and Park 2010). Element interactivity refers to the number of elements that must be processed simultaneously in the learner’s working memory while processing a task (Chandler and Sweller 1996). The learner’s existing knowledge is pivotal because new data can be connected to current schemas (Gerjets et al., 2004). Consequently, the processed information is more coherent and organized, preventing learners from managing unrelated items in their working memory. Extrinsic Load arises from a learning experience that is needlessly challenging or muddled. Mental resources spent on processes not directly related to the task, such as hunting for or suppressing information, are termed extraneous processes (Klepsch et al., 2017). Element interactivity, the presentation style of information, and required learning activities all contribute to cognitive load (Paas, 2013). For instance, giving data to participants using a countdown could lead to divided attention (Mayer and Moreno, 2010) because they must mentally combine distinct information sources that might have been more effectively integrated. Germane load pertains to the mental resources employed to process information, leading to the acquisition and automation of

schemata in long-term memory (Debie and Leemput 2014). It signifies the effort dedicated to integrating new knowledge into existing cognitive structures, enabling its transition into long-term storage. Sweller et al. (1998) introduced this concept, noting that certain instructional formats could elevate both cognitive load and learning simultaneously. Cognitive overload arises when the cumulative effect of intrinsic, extraneous, and germane loads becomes too much for an individual to handle. In such situations, an individual may find it challenging to absorb new information or make informed choices.

To sum up, ‘mental load’ serves as the basis of theories related to ‘load’, contributing to the measurement of cognitive load. Cognitive load is a multi-dimensional construct that includes aspects such as mental stress, cognitive effort, and working memory capacity (Paas et al., 2003). It symbolizes the load imposed on an individual’s cognitive system due to task requirements.

### **2.2.2 Cognitive Load Concept in Interpreting and Translation Studies**

CLT is considered as the most effective theoretical framework of cognitive processing in Interpreting and Translation Studies (Gile 2009; Seeber 2011; Chen 2017). Additionally, the manipulation of text difficulty and time pressure/delivery rate may affect an interpreter’s cognitive load, and cognitive load is modulated by the design of a training task. CLT is aimed at “supporting instructors in the development of novel instructional designs aligned within the limitations of the human cognitive architecture” (Orru and Longo, 2018, p24). The training content based on CLT can be helpful to the future interpretation training.

Cognitive load serves as the foundational principle for cognitive methodologies employed in the exploration of interpreting (Seeber 2011; Chen 2017). Specifically, within the interpreting context, cognitive load refers to the proportion of processing capacity expended on an interpreting task in a system that inherently has a finite capacity (Seeber 2013). Interpreting and Translation tasks are considered as a high cognitive demanded task, and different processing tasks are continuously vying for a limited pool of cognitive resources. Handling incongruities during interpreting can significantly increase the cognitive load, affecting memory storage and syntactic integration.

Seeber (2011, p. 187-189) defined cognitive load as the interference between various demands in “a real-time combination of a language comprehension and language production task”. These

demands can occur on both macrostructural (i.e., perceptual, cognitive, or response processing) and microstructural (i.e., processing of linguistic constituents) levels. Seeber (2013, p. 19) also borrowed Miller's (1956) idea and further defined cognitive load as "the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system". He postulates that cognitive load depends not only on the resources needed by individual tasks but also on the extent to which these tasks interfere with each other. Thus, tasks with similar processing structures tend to interfere more, thereby amplifying the cognitive load. Chen (2017, p. 643) further defined cognitive load in interpreting as "that portion of an interpreter's limited cognitive capacity devoted to performing an interpreting task in a certain environment." The concept of cognitive load offers a more comprehensive and broader perspective for understanding tasks like sight translation and simultaneous interpreting, which have significant cognitive demands.

### **2.2.3 Previous Literature on Cognitive Load in SI and ST**

The field of simultaneous interpreting and sight translation studies has seen burgeoning research on the concept of cognitive load. However, there has been a noticeable disparity in the volume of research dedicated to these two areas, with sight translation receiving greater scholarly attention. This section specifically delves into a review of selected literature on sight translation and simultaneous interpreting, with an emphasis on identifying notable research gaps. This literature review aims to situate our understanding within the current body of scholarly work. Furthermore, a table summarising the key information from the reviewed literature, especially aspects that have been insightful for my research, will be provided subsequently. After the Table 2, detailed explanations of each piece of literature will be given.

**Table 2 Research on Cognitive Load in Sight Translation and Simultaneous Interpreting**

| Scholar                      | Mode | Language            | Sample                                  | Independent variables                                                             | Dependent variables<br>(cognitive load indicator) | Highlights<br>(related to my research)                                                                                                                                                                        |
|------------------------------|------|---------------------|-----------------------------------------|-----------------------------------------------------------------------------------|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Shreve et al. (2010)         | ST   | Spanish – English   | 11 trainees with different degree level | syntactic complexity effect                                                       | eye movement patterns, errors in verbal output    | high demand of C effort during ST; syntactic effects may not increase cognitive load                                                                                                                          |
| Seeber and Kerzel (2012)     | SI   | German – English    | 10 professional interpreters            | verb-initial, verb-final, syntactically asymmetrical, amount of available context | mean pupil dilation                               | syntactic differences between languages cause higher cognitive load; simultaneous interpreters do not always work close to their cognitive limit; pupil size is a feasible index for measuring cognitive load |
| Plevoets and Defrancq (2016) | SI   | French – Dutch      | Dutch interpreting corpus               | delivery rate, lexical density, percentage of numerals, average sentence length   | the frequencies of <i>uh</i> 's and <i>uhm</i> 's | high delivery rate causes cognitive overload for interpreters                                                                                                                                                 |
| Mizuno (2017)                | SI   | Japanese – English, | 8 professional interpreters             | syntactic difference                                                              | chunks                                            | the concept of chunk plays a significant role in cognitive load                                                                                                                                               |

|                              |    |                    |              |                                                                                   |                                                                               |                                                                                                                                                                                                                |
|------------------------------|----|--------------------|--------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                              |    | English – Japanese |              |                                                                                   |                                                                               | study; interpreters use diverse syntactic restructuring strategies to avoid attention overload when interpreting between asymmetrical languages                                                                |
| Plevoets and Defrancq (2018) | SI | French – Dutch     | EPICG Corpus | delivery rate, lexical density, percentage of numerals, formulaicity, hesitations | the frequencies of <i>uh</i> 's and <i>uhm</i> 's                             | the significant impact of high delivery rate; the strategy of using more formulaic sequences may reduce cognitive load                                                                                         |
| Li (2019)                    | ST | English – Chinese  | 10 trainees  | syntactically non-complex sentences vs syntactically complex sentences            | average fixation duration, average fixation count, mean pupil size            | syntactic differences cause higher cognitive load during E-C sight translation; pupillometry is a sensitive indicator that needs to be treated with great caution                                              |
| Ma (2021)                    | ST | English – Chinese  | 23 trainees  | relative clauses, passive constructions, sentence context, discourse context      | dwell time, fixation count, regression path duration, first fixation duration | syntactic complexity increases cognitive load; wider contexts improved sight translation time efficiency, but they didn't reduce the effort in word-based processing; no correlations between reading span and |

|                      |    |                                       |                             |                                                                                   |                                                                               |                                                                                                                                                                                      |
|----------------------|----|---------------------------------------|-----------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                      |    |                                       |                             |                                                                                   |                                                                               | cognitive load in addressing syntactical complexity                                                                                                                                  |
| Shao and Chai (2021) | SI | English – Chinese                     | 9 professional interpreters | difficult sentences vs easy sentences                                             | chunks, interpreting accuracy, EVS length                                     | imported load don't impact simultaneous interpretation (SI) performance, but current load does; interpreters' cognitive load approach the capacity limit of working memory during SI |
| Ma et al. (2021)     | ST | Chinese – English                     | 25 trainees                 | discourse context vs single-sentence, word order asymmetry vs word order symmetry | dwell time, fixation count, first fixation duration, regression path duration | significant effect of word order asymmetry on overall cognitive load; the impact of context in counterbalancing the asymmetry effect was minimal;                                    |
| Ma and Li (2021)     | ST | English – Chinese                     | 23 trainees                 | chunking, reordering                                                              | dwell time, fixation count, rereading rate                                    | chunking was the primary strategy to deal with word-order asymmetry; reordered sentences cause greater cognitive load                                                                |
| Gumul (2021)         | SI | English – Polish,<br>Polish – English | 40 trainees                 | explicitation                                                                     | hesitation markers, false starts, unfilled anomalous pauses exceeding two     | high delivery rate and lexical density will increase cognitive load, retrospective comments; relying solely on retrospective                                                         |

|                 |    |                  |             |                                                                                                     |                                                             |                                                                                                                                                                                |
|-----------------|----|------------------|-------------|-----------------------------------------------------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 |    |                  |             |                                                                                                     | seconds, self-reports, retrospective comments               | comments may not accurately reflect interpreter's perceived cognitive load                                                                                                     |
| Gieshoff (2021) | SI | English – German | 14 trainees | see speaker's lip movements vs. cannot see lip movements, interfering noise vs no interfering noise | subjective reports, pupil size, interpreting accuracy score | pupil size changes likely relate more to the speaker's video dynamics than lip movements or a static face; self-reports and accuracy score may not match pupillometry results. |

The research on cognitive load in sight translation remains somewhat limited, with studies tending to concentrate on specific text-based factors, especially in Chinese-English language pair. The term text-based here refers to the linguistic features and characteristics inherent to the source text, which include, but are not limited to, aspects such as word-order asymmetry, syntactic complexity, and semantic density. These elements, when interacting with an interpreter's cognitive resources, significantly influence the quality and efficiency of sight translation. This emphasis on text-based factors may be attributed, in part, to the substantial differences between the English and Chinese language systems. These languages differ fundamentally not only in their writing systems but also in their phonology, grammar, syntax, and vocabulary. The stark contrast in these linguistic features adds an additional layer of complexity to the cognitive load of interpreters and translators, possibly explaining why research in this area has often focused on these text-based elements. For instance, my previous research during my master's program, as represented in the study by Li (2019), focused on the exploration of syntactic complexity's influence on cognitive load in the context of E-C sight translation. Employing eye-tracking technology as a central element, ten participants with four years of experience in sight translation were recruited to translate an English text into Chinese. The text was designed to include both syntactically complex sentences and syntactically non-complex sentences, serving as the independent variable in the study. In the study, the dependent variables (fixation count, fixation duration, pupil size) were metrics assumed to reflect cognitive load. The outcomes of this study implied that sight translating syntactically non-complex sentences required less cognitive exertion compared to the translation of syntactically complex sentences. Furthering this, a collection of studies released in 2021 by Ma and colleagues offered considerable insights into the influence of factors such as word order asymmetry and syntactic complexity on cognitive load during C-E sight translation. Despite being conducted with various collaborators, these studies consistently illustrated that an increase in syntactic complexity correspondingly raises the cognitive load during sight translation tasks. Notably, while chunking is predominantly employed to tackle word-order asymmetry in asymmetrical structures, it was observed that the reordering of sentences contributed to a notably higher cognitive strain (Ma et al., 2021; Ma, 2021; Ma and Li, 2021). Their results reveal the comparative cognitive efficiency of different strategies, elucidate the relationship between strategic choices and processing effort, and lend empirical support to the significance of incorporating chunking into C-E/E-C sight translation training programs.

Similarly focusing on syntactic effects in sight translation, the findings from Shreve et al. (2010) are particularly intriguing. In their experiment, Shreve and his colleagues asked participants to sight-translate two Spanish-English texts (A and B), each containing two paragraphs manipulated into non-complex and complex versions. Text A presented the non-complex version first, followed by the complex version. Conversely, Text B started with the complex version, then transitioned to the non-complex one. Shreve et al. (2010) initially hypothesized that disruptions due to syntactic manipulations would manifest in distinct eye movement patterns. However, when they compared the overall means for the syntactically complex and non-complex Areas of Interest (AOIs), they found no statistically significant differences, thus challenging their initial expectations. Upon further comparative analysis between Text A and Text B, they discovered a syntactic complexity effect present in text A but absent in text B. This means only the result from the study of Text A support the predicted syntactic effect on sight translation task. In Text B, Shreve et al. (2010) observed a noticeable decline in performance between the first and second paragraphs. Eye-tracking metrics also indicated increased cognitive effort as the task progressed. While one might consider fatigue as a potential cause, the authors dismissed this explanation, noting the brevity of the texts and the absence of time constraints. They posited that sight translation might require significant coordination effort, as per Gile's Effort Model, to manage the continual visual disruptions from the source text. This increase in effort over time suggests a potential visual interference effect. However, this raises a concern that the observed disparity in cognitive load between non-complex and complex texts in Text A might not stem from syntactic manipulation, but rather from the effects of visual interference. In other words, the impact of the syntactic effect on sight translation remains uncertain and merits further investigation.

Seeber and Kerzel (2012) employed a psycho-physiological method, utilizing pupillary responses, to gauge cognitive load during simultaneous interpreting in both single-sentence and discourse contexts. Their research also explored the cognitive implications of translating verb-final constructions into English, a verb-initial language. Results demonstrated a higher cognitive strain with verb-final structures, as evident from increased pupil dilation. This was particularly noticeable in the interpretation of German verb-final structures into strictly verb-initial English, resulting in a significant cognitive load spike. Seeber and Kerzel (2012) claimed that their findings contrary to the universalist theory of interpretation, which posits that the structural differences of the languages involved do not affect the process, however, their results suggest that these differences are indeed impactful. Moreover, the study did not support Gile's

(1995) ‘tightrope hypothesis’, which proposes that simultaneous interpreters constantly operate near the limit of their cognitive capacity. Instead, the results suggested that these capacity boundaries are only reached towards the end of the sentence during the interpretation of verb-final structures into verb-initial languages. Additionally, the absence of sudden or pronounced pupil contractions during the interpretation of the source sentences indicated that the participants did not suffer from cognitive overload. The observed peaks in cognitive load were relative rather than absolute maxima, suggesting a certain adaptability and flexibility in cognitive capacity during the simultaneous interpreting process. According to Seeber and Kerzel’s (2012) research, the physiological method offers an immediate and authentic reflection of the interpreter’s cognitive load, thereby providing a real-time assessment. This unique advantage amplifies the significance and utility of pupillometry as a tool in cognitive load research.

From another perspective, Gumul (2021) investigated the relation between explicitation and trainee interpreters’ cognitive load during simultaneous interpreting. In the source texts used for the experiment, the author intentionally included elements—such as delivery rate, lexical density, and proportion of numbers— which are hypothesized to increase cognitive load. Within the framework of the four approaches to cognitive load measurement outlined by Paas et. al (2003), this research applies the performance and subjective methods. The performance method is utilized in the product analysis, which involves a manual comparison of source and target texts, while the subjective method is implemented in the analysis of process data, which assesses retrospective participant protocols. In the product, the relationship between the frequency of explicating shifts and the surge in cognitive load is examined by pinpointing indicators of potential issues. These indicators are categorised into three types of disfluencies: hesitation markers, false starts, and unusually long pauses that last more than two seconds (a metric related to performance). Complementing the performance method, the subjective method of measuring cognitive load was employed through the analysis of process data. It entails reviewing participants’ self-reports for any mention of increased cognitive load or increased cognitive load. However, Gumul (2021) also asserted that even though the product analysis indicates a potential trend where explicitation could instigate cognitive load slightly more often than being a consequence of it, the statistical evidence does not demonstrate any significant distinction between the two. The actual cognitive load brought about by the simultaneous interpretation task and the experienced cognitive effort by the participants in the current study might possibly be heavier than what can be deduced from the visible disfluencies

because the problem indicators do not necessarily arise in the exact segments that turn out to be problematic. This may illustrate the limitations of the retrospective protocol and the importance of utilizing a triangulation approach in research.

Gieshoff (2021) conducted an empirical study to investigate cognitive load and mental effort during simultaneous interpreting with two variables (a) visual presentation (lip movements vs no lip movements), and (b) auditory presentation (noise vs no noise). The author also adopted three approaches from the cognitive load measurement methods outlined by Paas et. al (2003): the assessment of interpreting performance, the use of self-reports, and the application of psychophysiological measurements (pupillometry). In this study, Gieshoff (2021) delineated cognitive load as task demands perceived by the participants, shown through task difficulty ratings, and characterized mental effort as the observed response to these demands, evidenced by pupillometry and interpreting accuracy. The results from this study did not substantiate the facilitation hypothesis for lip movements. Interpretations were less aligned with the source speech in noisy conditions, and interpreters reported increased difficulty under these conditions. This indicates that overlaid noise heightens cognitive load during both listening and interpreting, leading to more mental effort expended during interpretation. In addition, contrary to expectations, no interaction between auditory and visual presentations was found in self-reports, interpreting accuracy, or pupillometry. Notably, the visual presentation had no significant impact on self-reports or interpreting accuracy, it did influence pupil sizes, with larger pupil diameters observed when lip movements were present. However, the author argued that observing larger pupil sizes in the presence of lip movements does not necessarily equate to lip movements inducing higher mental effort or cognitive load. This conclusion is drawn from the fact that lip movements did not result in any significant positive or negative changes in performance metrics or self-reports. Moreover, while noise noticeably influenced interpreting accuracy and self-reports, it did not have the same impact on pupil sizes. Therefore, it's suggested that the noticeable change in pupil sizes is more likely attributed to the dynamics of a speaker's video presentation, rather than the presence of lip movements or a static facial image. This intriguing finding suggests that self-reports and accuracy may not align with the results derived from pupillometry.

Furthermore, using the performance method, Plevoets and Defrancq conducted a series of studies in 2016 and 2018 to explore cognitive load during interpretation. They utilized the frequency of the speech disfluency 'uh(m)' as an indicator. In their research, they considered

four independent variables: delivery rate, lexical density, percentage of numerals, and average sentence length. Among these, the delivery rate of the source text emerged as a key predictor of ‘uh(m)’ occurrences in the interpreted text. While high delivery rates posed no issues for interpreters during production, they led to cognitive overload for interpreters when it came to processing incoming text. In contrast, high lexical density did not pose a problem during listening but was challenging during the production phase for interpreters. This pattern provided empirical support for theories of differential efforts (Gile 1995; Seeber 2011, 2013), suggesting that different tasks compete for the same cognitive resources. In addition, the frequency of ‘uh(m)’ in interpretations increased with the lexical density of the source text, while being inversely related to formulaicity in both the source and target text. Also, interpreters experienced lower cognitive load when they could use more formulaic sequences in the target language. The authors claimed that high delivery rates do not pose challenges for interpreters during production, they believe it leads to cognitive overload during the processing of incoming text. However, one must consider that processing and production typically occur simultaneously. Despite this reservation, the research of Plevoets and Defrancq (2016; 2018) highlights the significant impact of delivery rate on interpreters’ cognitive load, suggesting the need for further exploration in this area.

In addition to employing traditional methods of measuring cognitive load, several scholars have used alternative indicators as proxies for cognitive load. In 2017, Mizuno proposed a model for measuring cognitive load during simultaneous interpreting (SI) based on the amount of information held in focus of attention. This model equates cognitive load to the quantity of information held in an interpreter's working memory at any given moment. Arguing that Seeber's (2011) cognitive model did not adequately clarify the theoretical concepts of five demand vectors and interference, and had not been empirically tested, Mizuno presented a unique perspective. Drawing on Gile's (1995) Effort Model, Cowan's (1995) model of working memory, and Barrouillet and Camos's (2012) Time-Based Resource-Sharing model, Mizuno (2017) suggested that, assuming a high degree of automaticity in an interpreter, the cognitive load could be explained by the retention load within the focus of attention, where a unit of information or chunk is held. He summarised that “... the number of chunks held in the focus of attention can be deemed as a proxy of cognitive load derived from comprehension, language conversion, and production processes” (Mizuno, 2017, p. 13). The author claimed that this innovative approach operationalizes cognitive load in a manner that simplifies its measurement, providing a more streamlined means of assessing the cognitive demands on interpreters. The

findings partly substantiated Gile's 'tightrope hypothesis' and the concept of load transfer. Drawing from the same methodological framework as Mizuno (2017), Shao and Chai (2021) also evaluated cognitive load in their study as the volume of information contained within working memory, each distinct unit of which was identified as a chunk. This research involved a quasi-experiment with nine professional interpreters performing an English–Chinese simultaneous interpreting task. The results showed that imported load did not significantly affect SI performance, while the current load during an SI task was crucial. A negative correlation between local cognitive load and SI performance was found, with a notable decrease in performance observed when an interpreter's local cognitive load reached four chunks. This finding aligned with Cowan's (1995) suggestion that the focus of attention in working memory is limited in capacity, generally ranging from three to four chunks. A noteworthy outcome was the high frequency at which professional interpreters' cognitive load approached the capacity limit of working memory during SI, which supports Gile's (1995) 'tightrope hypothesis'. The study concludes that professional interpreters often operate close to cognitive saturation, even during relatively straightforward SI tasks. However, the applicability of the methods employed in Mizuno (2017) and Shao and Chai (2021), which use chunk as an indicator of cognitive load remains to be thoroughly tested. More empirical research is also needed, and it is still unclear whether these methods are suitable for all language pairs in Interpreting and Translation Studies.

In concluding the literature review on cognitive load research, it is clear that the cognitive load induced by delivery rate has been widely acknowledged in simultaneous interpretation, but it remains underexplored in sight translation. Second, the comparative impact of the delivery rate and other potentially problematic factors such as text difficulty on cognitive load and interpreting performance has barely been examined in English-Chinese language pairs. Moreover, the strategies employed by simultaneous interpretation and sight translation trainees in dealing with issues caused by external factors have not received extensive attention, let alone a comparative study of strategies under different interpreting modes. Third, currently, reaching a consensus on the validation of the Tightrope Hypothesis appears challenging. As Gile has noted, more empirical evidence is needed to modify the hypothesis. In light of the aforementioned three points, this research aims to provide empirical evidence in these areas, bridging gaps in our understanding and potentially guiding more effective training practices. Lastly, relying solely on one experimental method, whether it be analytical, subjective, performance, or psycho-physiological, may not yield conclusive results in measuring cognitive

load. This underscores the critical importance of employing mix-methods in cognitive load-related research to ensure comprehensive and robust findings. The following chapter will provide a detailed review of the four methods used to measure cognitive load, as well as their applications in the fields of interpreting and translation.

## **2.3 Methods of Measuring Cognitive Load**

Paas et al. (2003) and Schultheis and Jameson (2004) promote four approaches for assessing cognitive load: analytical method, subjective method, performance method and psychophysiological method. In this section, I will outline the application of these four measurement techniques utilized in Interpreting and Translation Studies. Subsequently, I will elaborate on the specific approach I have adopted within the broader method I have chosen.

### **2.3.1 Analytical Method**

#### **2.3.1.1 A Brief Overview of the Analytical Method**

Analytical method attempts to access cognitive load according to subjective data (e.g., expert opinion) and analytical data (e.g., mathematical models) (Paas et al. 2003). Typically, experts develop analytical and empirical frameworks based on current knowledge about the task to facilitate the measurement of cognitive load. For instance, Wickens (2002), Sweller (1988), Gile (2009), and Seeber (2011) have employed analytical methods in their studies. The analytical method, as Seeber (2013, p. 19) suggested, bypasses the need for “laborious empirical evaluations”, since both subjective and analytical information can be deliberated from a theoretical perspective. Daniel Gile’s Effort Model (1995) and Kilian Seeber’s Cognitive Load Model (2011) are the most-recognised analytical method for understanding cognitive load in Translation and Interpreting Studies. The former applies to different modes of translation and interpretation, including simultaneous interpreting and sight translation, while the latter is specifically designed for interpretation. In the present study, I borrowed the conceptual construct from Gile’s Effort Model as the theoretical framework to characterise the cognitive constraints of trainees while performing SI and ST tasks. A detailed explanation of this model will be provided in the following explanation.

### 2.3.1.2 The Tightrope Hypothesis and Effort Model

Gile's (1995/2009) Effort Model<sup>2</sup> is the most recognized analytical methods for understanding cognitive load in simultaneous interpreting and sight translation studies. Promoted by Gile (2009), the Tightrope Hypothesis posits that "interpreters tend to work close enough to cognitive saturation for many Errors, Omissions and/or Infelicities (EOIs) to occur not because the interpreters' insufficient knowledge of the working languages or the topics or insufficient technical skills but because attentional resources required to perform adequately were not available for a particular comprehension, memory storage or retrieval or production task at a time when they were needed" (2016, p. 9). To be specific, the Tightrope Hypothesis proposes that interpreters often operate near their cognitive limits, whether considering their overall processing capability or individual Efforts. This can be attributed either due to the cumulative demands of various efforts or sub-optimal resource allocation among them (Gile 2009). Contrary to the assumption that the available cognitive capacity is always sufficient for interpreting needs, the Tightrope Hypothesis suggests that failures in interpreting may sometimes arise from an ongoing tension between cognitive resources availability and demand, rather than merely from linguistic shortcomings or errors.

The Tightrope Hypothesis also offers a potential explanation for why certain interpreters or translators frequently encounter difficulties. These challenges may not stem from a lack of requisite knowledge but rather from the speed or density of the speeches they are handling; in essence, they might not possess the ability to process the information quickly enough (Gile 2016). Aligning with the variables of delivery rate and text difficulty explored in my research, this hypothesis may provide insight into why some unexpected errors and omissions occur, suggesting they are more attributable to limitations in processing capacity than to other factors.

Daniel Gile also created the Effort Model, drawing upon the idea of processing capacity rooted in cognitive psychology, to explain the interaction of various cognitive operations related to different type of interpreting and translation tasks. These sets of operations are categorized into Efforts, all competing for a limited processing capacity. When the total required processing power surpasses the available capacity, it results in a cognitive imbalance among the Efforts, leading to various types of errors. This "conceptual framework" (Gile, 2008, p. 62) bears

---

<sup>2</sup> Professor Daniel Gile regularly updates clarifications of the Effort Model and Gravitational Model on his personal research platform (ResearchGate). As of the submission of this dissertation, the most recent update was in January 2023.

similarity to Kahneman's (1973) single resource theory, which suggests a singular, limited pool of processing capacity that drives all cognitive tasks. The most notable feature of this model is that it simplifies the intricate different type of interpreting and translation modes into an uncomplicated mathematical formula (Seeber 2013). It should be noted that, the complete significance of the Effort Models in elucidating interpreting challenges truly emerges when paired with the Tightrope Hypothesis (Gile 2009).

According to Gile, "the efforts involved in SI can be defined as L (Listening and Analysis Effort: includes all reception and comprehension-oriented activities) + M (Memory Effort: interpreters store the receptive information in their short-term memory at the duration that the source language is heard and the completion of its production) + P (Production Effort: all operations from the mental representation of meaning to its real production in the target language) + C (Coordination Effort: cognitive effort demanded for managing the other three efforts)" (Li, 2019, p. 30). The Effort Model for SI is as follows:

$$SI = L + M + P + C^3$$

The Listening and Analysis Effort, as defined by Gile (2009, p. 160), "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". Interpreters are not merely tasked with constantly listening to the source speech, but also with decoding the message embedded in the source text. The process of comprehension during interpretation is not automatic. Rather, it actively involves some form of semantic representation of the source speech content, including an analysis of plausibility and likely some degree of anticipation (Gile 2009). Experimental studies (Gerver 1975; Lambert 1988), which monitor real-time inference and anticipation by interpreters, also have demonstrated that comprehension extends beyond merely the process of speech recognition. Within the framework of the listening and analysis effort, certain elements can result in an increased cognitive burden, particularly in challenging speech scenarios. These include complex syntax, rapid speech pace, accents, intonation etc (Derwing and Munro 2009; Cheung 2013). When these processing demands surpass the interpreter's cognitive capacity, it could result in a decline in their performance.

---

<sup>3</sup> The symbol "=" should be interpreted as "consists of" (Gile, 2009, p.168), while "+" signifies the combination of efforts.

The (short-term) Memory Effort “as it was conceptualized for the students referred to their experience of storing in memory over a short period (up to a few seconds) information captured from the source speech and then retrieving it for the purposes of target speech production” (Gile, 2021, p. 142). During interpreting, continuous short-term memory operations last for brief moments. Some operations arise from the time gap between hearing speech sounds and interpreting them, requiring phonetic segments to be analysed in memory. Others relate to the time required for speech production, maintaining the message’s idea in memory. If a speaker’s speech is unclear due to various factors, such as logic, density, structure, or accent, an interpreter might delay reformulation or note-taking to better comprehend and reformulate (Gile 2009). Language-specific factors also necessitate short-term memory operations. Gile (2016) in the Clarification and Update of Effort Models further emphasized that the Memory Effort should not be conflated with Working Memory (WM). While WM plays a role in both the Reception and Production Efforts, Memory Effort embodies a functional perspective, often intertwined with tactical or strategic considerations, such as whether the interpreter should wait before proceeding.

Production Effort during SI 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” (Gile, 2009, p. 163). During interpreting task, output is a mental process that bridges the representation of information in the brain to its verbal expression. Production Effort is not merely the act of verbalizing a message. It is a comprehensive process that demands both linguistic prowess and metacognitive skills. The interpreter, in this context, is not just a linguistic translator but a mindful communicator, always attuned to the accuracy and fluency of their own output. This means that as interpreters navigate the complexities of language conversion, they are also continually self-monitoring, ensuring that the spoken output aligns with the intended message.

The Coordination Effort encompasses the cognitive exertion required to manage and coordinate the other three efforts in the interpreting process (Gile 2009). The Coordination Effort directs students to the realization that they must adjust their attention according to the specific speech segments they are processing. While the inclusion of the Coordination Effort to the three core Efforts was inspired by the cognitive psychology concept of executive functions, its role in the Effort Model is distinct. Like Listening and Analysis Effort, Memory Effort, and Production

Effort, the Coordination Effort is functional in nature. Essentially, it helps allocate appropriate attention to specific tasks, such as listening or production, ensuring that sufficient resources are assigned without compromising another task. This means making strategic decisions, such as prioritizing comprehension of incoming speech rather than perfecting the target output. Additionally, it requires the interpreter to decide when to pause briefly before interpreting or when to closely follow the speaker's pace. Such coordination is crucial to ensure effective interpretation and mitigate cognitive overload.

On the other hand, in sight translation, the Listening and Analysis Effort in simultaneous interpreting transforms into the Reading Effort (R), while Production Effort and Coordination Effort remains consistent. The Effort Model for ST is as follows:

$$ST = R + M + P + C$$

In this context, the Reading Effort pertains to the visual acquisition of source information through reading. The Production Effort, similar to that in the SI model, bridges the visual information with oral expression. Sight translators constantly self-monitor during this process, dealing with the complexities of language conversion to ensure their verbal output aligns with the source message. The Coordination Effort, meanwhile, coordinates and manages the other efforts. For instance, it aids sight translators in deciding at which point in a read sentence they should commence their oral rendition. It should be noted that while the continuous visual presence of the source text, along with its self-paced nature, minimizes the risk of overloading the working memory, thereby suggesting a reduced cognitive load. However, as Gile (2009) highlighted, Production Effort is very difficult because the continuous visual engagement with the source text during sight translation can elevate the risks of linguistic interference.

In both SI and ST, there can be instances where one, two, or all three core Efforts are active (Gile 2009). For example, an interpreter might be delivering a previously prepared segment during a speaker's pause or after reading a segment of text, activating only one Effort. At other moments, they could be listening or reading and retaining information without initiating speech or planning its reformulation, indicating two active Efforts. Moreover, interpreters may simultaneously listen (or read in ST) and speak (Gerver 1975; Čenková 1988), thus engaging all three Efforts concurrently. Operational processing capacity requirements during SI and ST can be showed as follows (Gile 2009):

$$TR_{SI} = LR + MR + PR + CR$$

$$TR_{ST} = RR + MR + PR + CR$$

Where:

- $TR_{SI}$  and  $TR_{ST}$  represent the total processing capacity requirements for SI and ST modes, respectively.
- LR and RR stand for processing capacity requirements for Listening and Reading efforts, respectively.
- MR, PR, and CR stand for processing capacity requirements for Memory, Production, and Coordination efforts, respectively.

For successful SI and ST, the following five conditions must be consistently met:

$$TR \leq TA$$

$$LR \leq LA \text{ or } RR \leq RA$$

$$MR \leq MA$$

$$PR \leq PA$$

$$CR \leq CA$$

Where:

- TA represent the total available processing capacity, and the “total processing capacity requirements should not exceed the interpreter’s total available processing capacity” (Gile, 2009, p. 170).
- LA, RA, MA, PA, and CA stand for the processing capacity available for Listening, Reading, Memory, Production and Coordination efforts, respectively.

These conditions represent the equilibrium between the processing capacities required for each effort (on the left side of the inequality) and the actual available capacities (on the right side). When this equilibrium is disturbed, it means that one or several of the efforts are overextended. If the processing demands exceed the available capacities for any of these efforts, it can have detrimental effects on the SI and ST tasks. Specifically: “Incomplete/incorrect comprehension of the source speech; and/or Incorrect/clumsy target speech; and/or Incomplete/incorrect storage/retrieval of information from short-term memory; and/or Slowing down of one or several Efforts’ performance and chain reactions. All of these can result in Errors, Omissions and/or Infelicities (EOIs)” (Gile, 2016, p. 8).

The Effort Models dissect the intricate tasks of SI and ST into distinct efforts, simplifying the study and understanding of each individual component. By studying the cognitive constraints inherent in SI and ST, the Effort Models assist trainees, trainers, and researchers in identifying potential challenges and discerning the causes of unsatisfactory outputs. Moreover, it provides interpreters with insights into efficiently allocating cognitive resources, especially during challenging scenarios such as source language with a high delivery rate or source language with a high difficulty level.

### **2.3.2 Subjective Method**

#### **2.3.2.1 A Brief Overview of the Subjective Method**

Subjective method employs rating scales, requiring subjects to self-report their perceived cognitive load. Such methods are used to quantify phenomena that are typically viewed as challenging to assess objectively (Seeber 2013). Rating scale methods operate based on the assumption that individuals can reflect on their cognitive activities and report the mental effort they exerted (Pass et. al 2003). The majority of subjective evaluations are multidimensional, evaluating clusters of related variables. Although the subjective method has been criticized for its strong reliance on retrospection and the potential biases of delayed self-reporting, it remains a commonly used approach to measure cognitive load. For example, Hart and Staveland (1988), Ericsson and Simon (1984), Lamberger-Felber (2001) have adopted this method in their research.

#### **2.3.2.2 Task Load Index**

In the present study, I used National Aeronautics and Space Administration Task Load Index (NASA-TLX) to measure participants perceived cognitive load. It is a widely used self-assessment tool designed to measure the perceived workload or cognitive load associated with a specific task (Hart and Staveland 1988). Developed by the Human Performance Group at NASA's Ames Research Centre, the NASA-TLX has been employed across various domains, including interpreting studies, to assess the subjective workload experienced by individuals. Several scholars have demonstrated the feasibility of using NASA-TLX in English-Chinese interpreting and translation research. For instance, Chen (2017) and Wang (2017) highlighted its applicability for measuring cognitive load in interpreting, while Weng (2022) and colleagues employed this method in their study to explore time pressure in translation study. The NASA-TLX consists of six dimensions (NASA ARC 1980): Mental Demand (the amount of mental

effort required to perform the task, including cognitive processes such as thinking, decision-making, and problem-solving); Physical Demand (the degree of physical effort or activity required to complete the task); Temporal Demand (the perceived time pressure or the need to perform the task quickly); Performance (the individual's perception of their success or effectiveness in completing the task); Effort (the amount of overall effort, combining both mental and physical aspects, exerted to achieve the level of performance); Frustration Level (the extent of annoyance, irritation, or stress experienced while performing the task. Participants rate each dimension, with higher scores indicating a higher perceived workload.

### **2.3.3 Performance Method**

Performance method focus on the participant's performance quality in the empirical study, it "enables feedback and response from the interpreting output, which might contribute to 'making sense' of the data" (Yang, 2019, p. 242). Typical performance variables are reaction time, accuracy, and error rate (Paas et. al, 2003, p. 66). When cognitive load exceeds an interpreter's capacity, performance decreases, making it a sensitive indicator of load changes; in these situations, a performance score can reflect the overall cognitive load of a task (Chen 2017). In interpreting and translation education, performance scores also used to indicate the complexity of training tasks, typically complemented by subjective reports or expert insights.

For a long time, interpreting studies have employed performance methods to gauge cognitive load by assessing accuracy (Seeber 2013; Chen 2017). Beyond measuring accuracy, other performance metrics, like the span of target speech and error counts have also been extensively researched. For instance, Oléron and Nanpon's (1965) quantification of the Ear to Voice Span (EVS) offer insights into specific aspects of interpreting. Barik (1973) analysed errors and omissions in simultaneous interpreting.

This study will choose to assess the accuracy of trainees to collect performance-based data. However, since this section involves how to define the concept of accuracy and how to establish an evaluation criterion for this study, I will delve into these aspects in detail in the Methodology Chapter.

### **2.3.4 Psycho-Physiological Method**

#### **2.3.4.1 A Brief Overview of the Psycho-physiological Method**

Physiological approaches “are based on the assumption that changes in cognitive functioning are reflected by physiological variables” (Paas et. al, 2003, p. 66). These techniques include “the measurement of cardiac, hematic, electro-dermal, ocular, muscular and cerebral responses” (Seeber, 2013, p. 25). In other words, these approaches contain measures of heart activity, brain activity and eye activity (Paas et. al 2003). Several of these approaches offer continuous monitoring, enabling real-time event analysis. This is particularly significant for language processing because it is believed that cognitive load varies intermittently, and the specific tasks involved in language processing occur swiftly, sometimes within mere fractions of seconds (Mitchell 2004). Psycho-physiological metrics are optimal for showcasing intricate load trends and patterns, such as instantaneous, peak, average, and cumulative load.

With the advancement of cognitive research, eye-tracking technology has been increasingly utilized in Translation and Interpreting Studies. By observing eye movements, insights can be gleaned into both the physical responses and cognitive processes of the translator while engaged in a cognitive task (Richardson et al. 2004). Among eye-based measurement techniques, the long-established method of pupillometry offers better temporal precision, is less intrusive, and more cost-effective. Additionally, as Translation and Interpreting Studies take a psycho-affective turn, research is anticipated to increasingly emphasize neurophysiological dimensions, integrating neuroimaging studies alongside behavioural investigations (Alves and Albir 2017). Thus, scholars are increasingly exploring the feasibility and applicability of neurophysiological measures such as electroencephalogram (EEG), galvanic skin response (GSR), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) in Translation and Interpreting Process Studies. In this study, I employed both pupillometry and GSR as psycho-physiological measurement tools. The following sections showed the application of these two techniques.

#### **2.3.4.2 The Application of Pupillometry**

People always say, “eyes are the window to the soul.” This section provides a comprehensive introduction to the human pupil and its relationship with cognitive load, specifically focusing on changes in pupil size as an indicator of cognitive load fluctuations. Besides, I also explored its applications in the fields of translation and interpreting.

The human eye is a complex and sophisticated organ that plays a crucial role in our perception of the world around us. Observations as early as the mid-18th century noted that pupil dilation occurred independent of changes in illumination, suggesting that there is more to it than mere light response (Loewenfeld 1958). This discovery paved the way for subsequent research. It was found that the size and response dynamics of pupils can be utilized to deduce a range of psychological and physiological conditions since pupils can react to different stimuli such as alterations in light levels, cognitive load, and emotional stimulation (Adhikari et al. 2015, Binda 2015).

The pupil, a central component of the human visual system, is the small, black circular opening in the middle of the iris, the coloured part of the eye (Mathôt 2018). Serving as a primary filter for visual information, the pupil can vary in size: from 1.5 to 2 mm during full constriction to between 7.5 and 8 mm during full dilation. As a result, there is a factor of thirty-six difference between the area of the pupil ( $\pi r^2$ ) and the amount of light admitted to the eye (Alexandridis 1985). A primary function of the pupil is to regulate the amount of light that enters the eye, determine the depth of field, and focus the image formed on the retina. The diameter of the pupil is controlled by two sets of muscles: the radial dilator muscles and the circular sphincter muscles. The radial dilator muscles are in the outer part of the iris and function to pull the iris outward, thereby enlarging the pupil; on the other hand, the circular sphincter muscles are in the inner part of the iris and function to constrict the pupil (Ebitz and Moore 2017). These changes have an impact on both the quality of our perceptions of the visual world and the amount of light that reaches the retina. The regulation of pupil size is a complex process that involves the interaction of multiple neural and physiological systems. No other smooth muscle in the human body can shorten its fibres by as much as eighty-seven percent during maximal constriction, which is a characteristic of the sphincter pupillae (Alexandridis 1985). The diameter of the pupil is controlled by the autonomic nervous system, which consists of the sympathetic and parasympathetic branches. As the diameter of the pupil changes, it affects the depth of field and focus of the image that is formed on the retina. The research on the change of pupil size provides insight into the complex interactions between the neural and physiological systems.

Having gained a comprehensive understanding of individual's pupil, to further contextualize our methodological approach, it is imperative to further elaborate on the interplay between

pupil dynamics and cognitive load. As previously mentioned, the change in pupil size is determined not only by physiological factors but also by its response to various conditions, including emotional states and cognitive load. Individual's pupil size increases with an increase in cognitive load because the sympathetic nervous system is activated under increased cognitive load, leading to pupil dilation as a physiological response. The sympathetic nervous system is responsible for the 'fight or flight' response. When cognitive load increases, the sympathetic nervous system is activated, leading to the release of norepinephrine which causes contraction of the iris muscles, and thus dilation of the pupils. These facts suggest that pupil dilation can serve as an indicator of cognitive load levels. Various studies have reinforced the utility of pupillometry as a robust tool in assessing cognitive load and processes. Hess and Polt (1960) made early strides in this domain by illustrating that pupil size fluctuates in relation to the interest value of visual stimuli, emphasizing a tangible link between pupil dilation and cognitive processes. Following this trajectory, Kahneman and Beatty (1966) presented evidence that pupil size expanded stepwise with each additional item participants were tasked with memorizing. This pattern indicates that pupil diameter can be a reliable measure of cognitive load, specifically related to memory tasks. Such observations gave rise to the cognitive effort hypothesis, asserting that the degree of pupil dilation directly correlates with the cognitive load a person experiences. Beatty (1982) consolidated evidence linking task-evoked pupillary responses to cognitive processing load. Similarly, the relationship between pupil size, mental effort, and cognitive load was affirmed by Tammola and Niemi (1986). Granholm and Steinhauer (2004) further validated the cognitive effort hypothesis through their work with pupillometry. Extending the narrative, Ebitz and Moore (2019) proposed that cognition might modulate pupil behaviour to selectively filter visual information. Collectively, these studies emphasize the consensus in the scientific community regarding the correlation between pupil dilation and cognitive demands, reinforcing the methodological rationale for using pupillometry in such assessments. The pupil dilation in response to cognitive load usually seems to begin after 300–500ms (Hoeks and Levelt 1993). "Unlike heart-rate variability and other physiological measures, the cognitive pupillary response seems a highly sensitive instrument for tracking fluctuating levels of cognitive load" (Pass et. al, 2003, p. 66).

Pupil size measurements have been a significant tool in Interpreting and Translation Studies to gauge cognitive load. Tammola and Niemi (1986) were pioneers in this domain, revealing that sentences needing syntactic restructuring led to increased cognitive load, as evidenced by pupil dilation. A subsequent study by Tammola and Hyönä (1990) demonstrated a progressive

increase in cognitive load across different tasks - from listening to simultaneous interpreting. Furthermore, Hyönä et al. (1995) confirmed the sensitivity of pupil dilation as an index of cognitive load when interpreting different levels of translatability in words. Su and Li (2019) emphasized the challenges in sight translation directionality, noting greater cognitive strain when translating into a foreign language and particularly with complex structures. Kumcu (2011) explored reading patterns during simultaneous interpreting and highlighted the distinct nature of preparatory reading versus on-the-fly reading during interpretation. Gieshoff (2021) conducted an experiment comparing cognitive load in interpreter trainees with different types of inputs (audio-visual vs. static image). Delving into syntactic processing, Seeber and Kerzel (2012) employed eye-tracking and pupillometry to study the impact of verb placement in sentences during interpreting. Building on similar themes, Korpál (2012) found that German texts, due to their unique syntactic structures, posed more challenges than English texts. To summarise, these studies underscore the invaluable role of pupillometry in understanding cognitive load and processing dynamics in the realm of interpreting and translation.

#### **2.3.4.3 The Application of GSR**

This section offers a detailed introduction to GSR and its application in cognitive load research. Galvanic skin response is the continuous changes in the electrical properties of the human skin, also known as the psychogalvanic reflex (PGR) or electrodermal response (EDR), refers to the variation in the electrical conductance of the skin due to sweat gland activity in response to physiological arousal (McCleary 1950; Montagu and Coles 1966; Sharma et al. 2016). The term Galvanic skin response is associated with the Italian physician and physicist Luigi Galvani, who first discovered the relationship between electricity and the nervous system in the 18th century (Boller et al. 1989). Subsequently, GSR as the changes of skin conductance formally introduced by Charles Féré in 1888 and then widely used in psychophysiological research as a non-invasive measure of the autonomic nervous system's response to various stimuli. As the primary interface between body organisms and outside environment, our skin can tell us a lot of information about our feelings of different types of stimuli. Initially, muscular theory, vascular theory and secretory theory (sweat-gland related) were the three main ideas that might support the physiological basis of the GSR (McCleary 1950). Decades of research support the notion that GSR is mediated only by the sympathetic cholinergic nerve provide to human skin, which is closely related to sweat-gland activity (Lader and Montagu 1962; Montagu and Coles 1966). The human body has about three million sweat glands, they secrete moisture through

pores towards the skin surface when they are triggered and turn more active. Many scholars (Golla 1921; Merwe and Theron 1947; Theron 1948; Mundy-Castle and McKiever 1953) proved that GSR is linked with human emotion as it changes in the skin and subcutaneous tissue when elicited by complicated stimuli or when human affective state changes, and the rate of changes of human finger volume is also positively correlated with emotional arousal. This is why the GSR transducer is placed on the finger of our hand because the phenomena showed by fingers can tell us something useful. One benefit of GSR is when compared with other neurological methods such as fMRI or EEG, GSR allows participants to wear it more comfortably and requires a very brief preparation and calibration time. Nowadays, GSR is regarded as the most effective and popular method for investigating human psychophysiological phenomena (Boucsein 2012).

Numerous researchers have confirmed that GSR can serve as an indicator of cognitive load. A team from NICTA (Australia's Information and Communications Technology Research Centre of Excellence) employed GSR in 2007 to objectively assess users' arousal levels when using unimodal and multimodal versions of the same interface. Shi et al. (2007, p. 2851) found that "users' GSR readings significantly increase when the task cognitive load level rises. Additionally, users' GSR readings were lower when using a multimodal interface compared to a unimodal one. A detailed examination of the GSR data, aligned with multimodal data annotation, offered insightful explanations for peaks in the GSR data, correlating them with specific sub-task user events." This suggests that GSR can effectively measure users' cognitive load with a suitable level of granularity. In 2012, a team from NICTA used GSR to measure cognitive load during arithmetic and reading tasks. Nourbakhsh et al. (2012) examined both temporal and spectral features of GSR across different task difficulty levels, emphasizing the significance of the spectral features. Expanding on this, Nourbakhsh et al. (2013) further analyzed attributes of GSR and eye blink signals in classifying cognitive workload levels. Their experiment, which involved arithmetic tasks of varying difficulty, applied classification algorithms to the collected GSR and blink data. The results indicated that combining features from both modalities enhanced cognitive load classification accuracy, underscoring the utility of merging skin electrical conductance with eye activity metrics. Additionally, Ferreira et al. (2014) integrated GSR with EEG (electroencephalogram) to investigate perceptual speed and visio-spatial cognitive processing capabilities. Their models accurately differentiated between low and high cognitive loads for both younger and older adults. The psycho-physiological data they gathered indicated the real-time cognitive load levels experienced by these two age groups,

further attesting to the feasibility of this method for cognitive load assessment. Khawaji (2015) employed GSR to gauge trust and cognitive load in text-chat environments, where participants exchanged messages through software on smartphones and computers. The study found significant GSR variations influenced by both trust and cognitive load, suggesting GSR's potential as an effective cognitive load measurement tool. Lochner et al. (2019) explored trust and cognitive load during semi-automated unmanned aerial vehicle operations using GSR. Their research concluded that trust states could be classified based solely on GSR data, presenting a novel method that could foster understanding of the interplay between automation, user trust, and cognitive load. In light of the discussed literature, GSR emerges as a potentially valuable tool for assessing cognitive load.

## **2.4 Experimental Variables**

In this section, we will examine the roles of text difficulty and delivery rate as independent variables, with working memory capacity considered as covariate variable.

### **2.4.1 A Brief Description of the Input Variables**

The intricate coordination of comprehension and production in both SI and ST processes, influenced by several 'external' factors, primarily pertains to the nature of the source message acting as the immediate input for the interpreter's cognitive operations (Pöchhacker 2016). The input that the interpreter must handle does not solely consist of the content and form of the source language, but it is also influenced by the working conditions under which it is received (Mead 2015). As Pöchhacker (2016, p. 122) mentioned in the seminal guidebook *Introducing Interpreting Studies* – "Prior to source-text processing as such is the issue of acoustic and/or visual access and perception".

The interest of input factor can be traced back to early empirical studies when experts viewed interpreting as a unique form of real-time language processing. Gile (1995) highlighted three main areas that might cause cognitive overload during interpreting: 1) the 'problem trigger', which involves objective factors related to the linguistic, semantic, and physical attributes of the source speech; 2) the environment, covering elements like noise, stress, and preparation; 3) the interpreter-specific attributes, such as working memory, attention, and cognitive processing capability. Setton (1999) identified three categories of input variables: 1) the 'language in text',

which relates to the style, presentation, and delivery of speech; 2) the subject, which focuses on interpreter-related factors like competence and motivation; 3) the environment, focusing on factors like the audience's characteristics and technical conditions. Kraviarová (2013) suggested that tasks are influenced by multiple factors, including the source text, the speaker, the interpreter, and environmental conditions. Kraviarová (2013) detailed five types of input variables that could create challenges: 1) sender-related issues like accent, speech rate, and intonation; 2) message-related challenges such as terminology and proper names; 3) semiological problems, including metaphors and humour; 4) interpreter-associated issues, such as experience and background; 5) technical complications, like equipment and visual settings. Mead (2015) posited that the source language input comprises the content and form of the speech and the operational context. The content entails aspects like the density of information and terminology frequency. The form captures the complexity of syntactics, source text's sentence length, lexical density, and delivery rate. Additional influential variables, include technology, sound quality, equipment, and visual accessibility to speakers and materials. Pöchhacker (2016) also distinguished four groups of input variables: 1) sound and vision, such as technical equipment related to acoustic quality and visual access; 2) accent and intonation, such as speaker's pronunciation; 3) speech and mode of delivery, focusing on delivery rate, pause criterion or prosodic patterns; 4) source-text complexity, focusing on lexical level of the source text, text difficulties, text type and propositional density.

A significant portion of the referenced literature appears to agree that factors related to the source text, source language delivery, and the environment are crucial elements of source input variables.

#### **2.4.2 Delivery Rate as An Input Variable**

In simultaneous interpreting, according to Pöchhacker (2016, p. 124), "the speed of message delivery, also referred to as speech rate, 'input rate,' 'presentation rate' or 'delivery rate,' stands out as a key input variable in its own right." The form of a source message is not solely determined by its lexical and syntactic features; it is crucial to consider elements like the source language's delivery speed (Gerver 1969; Gile 1995; Shlesinger 2003; Riccardi 2015). The delivery rate of source language emerges as a pivotal factor, influencing fluency and determining the amount of information an interpreter processes within a specific timeframe

(Mead 2015). Gile also considered high delivery rate as the “cognitive problem trigger” (2009, p. 192).

In simultaneous interpreting, delivery rate “also referred to as speech tempo or speed, measures the words (or syllables) per minute (or second) of a speaker’s actual output, including PAUSES” (Riccardi, 2015, p. 397-398). Goldman-Eisler (1968) contended that pauses play a pivotal role in variations in speech rate. As the duration of speech delivery extends, discrepancies in individual utterances tend to diminish. The input rate suggested for SI by some scholars (e.g., Gerver 1976; Seleskovitch 1978; Lederer 1981) is the ranges between 95 and 120 words per minute. On the other hand, based on findings from Li (2010) and International Association of Conference Interpreters (AIIC) (1999), that rates between 100 and 120 words per minute are ideal when interpreting from English, with rates exceeding 120 wpm being especially problematic when dealing with content-heavy speeches or read-out texts. According to this, it is evident that a rate exceeding 120 words per minute is considered fast, while a pace between 100 and 120 words per minute is deemed moderate.

In the realm of interpretation and translation, a definitive standard for delivery rate in sight translation remains elusive. In my study, aiming for a balanced comparison between sight translation and simultaneous interpreting, I have opted to base our speed criteria on the established standards for simultaneous interpreting. Further specifics will be elucidated in the Methodology Chapter.

#### **2.4.3 Text difficulty as An Input Variable**

On the other hand, turning our attention to the textual-related variable. As Pöchhacker (2016, p. 125) argued, “One of the most difficult parameters of input load is the information content of the source message”. The level of text difficulty poses a significant challenge for interpreters. At the lexical dimension, studies have explored various elements relevant to interpreting. These aspects include word frequency, lexical variability, and specialized terminology, non-redundant terms like proper names and numbers, false cognates, culture-bound terms or humorous (Gile 1984, Viaggio 1995, Shlesinger 2000, Pöchhacker 2007). Despite the importance of these research, the overall difficulty of a text cannot be solely attributed to challenges at the lexical dimension. At the sentence level, studies focus on the relationship between propositions and their syntactic context, focusing on aspects like clause density,

embedding, and overarching textual structures (Dillinger 1994; Tommola and Helevä 1998). In addition, the application of traditional readability measures, like Flesch–Kincaid or SMOG, based on sentence length and word length also used to investigate text difficulty. While this method might not fully capture the difficulty of a text, it can still provide some insight into its potential difficulty.

When using text difficulty as the independent variable in a study, a pressing question arises: how do we measure the difficulty of a text? In the following part, I will review the methods of measuring text difficulty and revisit the approaches others have employed when considering text difficulty as an input variable.

Since one of the aims of this study is to observe how the difficulty level of a source text impacts the trainees' cognitive load, a critical issue that arises is the determination of a text's difficulty level. This is an essential procedure that helps to limit potential interference from other variables in the research (Lamberger-Felber 2001). Prior research has put forward numerous metrics for assessing the difficulty of source texts in the fields of Translation and Interpretation Studies. These metrics can be broadly categorized into two types: those that provide objective measurements and those that offer subjective evaluations (Yang 2019).

From the perspective of objective measurements, also known as quantitative measurements, Dam (2001) proposed the following indicators for interpreting experimental context: the number of specialised terms, numbers, average sentence length, clause length, words per clause, and the average rate of speech. Jensen (2009) focused on the concept of 'complexity', and suggested that the complexity inherent in standard texts not only serves as a strong indicator of difficulty but also offers a more objective approximation of relative text difficulty, drawing upon one or more factual criteria. Jensen (2009) further proposed three quantitative measures for evaluating source complexity, namely readability, word frequency and the non-literal expressions. This methodology was later adopted by Hvelplund (2011) in his cognitive research. Hild (2011) used three measurements for a SI study, including lexical parameters, syntactic measures, and discourse parameters. Pavlović and Jensen (2009) employed the Lix formula (based on word length and sentence length) and SMOG scale (based on syllable count and sentence length) in their comparative study on eye tracking in translation directionality. Sjørup (2013) utilized the Flesch Reading Ease index to determine the difficulty of texts in the context of a translation experiment. When it came to interpreting and translation research involving

Chinese-English language pairs, Liu and Chiu (2009) adopted readability formula, information density and new concept density to evaluate the level of text difficulty. Wang (2017) in a study of linguistic metaphors in English-Chinese translation assessed text difficulty on three distinct levels: the textual level (utilizing readability), the sentence level (using the number of sentences and average sentence length), and the word level (employing word frequency and word length). Ma (2019) utilized word frequency as a standard for quantitative measurement in a sight translation study.

When considering subjective measures - also referred to as qualitative measures - various researchers have chosen to utilize these methods in their studies. It is important to note that in recent years, some researchers have not been solely dependent on either quantitative or qualitative measures to gauge the complexity of the source text. More commonly, a blend of both these measures is employed to offer a more holistic view of source text difficulty and to facilitate cross-checking of data (Yang 2019). Chall et al. (1996) put forth a method that is dependent on the general impression of textual characteristics for the qualitative evaluation of text difficulty. Dam (2001) also suggested the feasibility of presenting two different texts to separate groups of interpreting students and assessing which text is simpler based on their subjective judgments and evaluations. Liu and Chiu (2009) incorporated the use of expert opinion as a technique to evaluate the difficulty of source material in their study. Hvelplund (2011) invited three evaluators to complete a questionnaire designed to rate the comprehensibility, coherence, and grammatical correctness of the source texts. In Chang's (2011) study, ten bilingual individuals were asked to evaluate both the comprehensibility and readability of the source text, using a grading system ranging from 1 (indicating very easy) to 5 (indicating very difficult). In addition to this, two professional translation teachers were solicited to appraise the translatability of the same text, utilizing the same scoring system. Sjørup (2013) additionally sought the input of an evaluator to provide insights on the readability of the source text and its level of comprehensibility. Ma (2019) also gathered a group of experts to rate the text's readability on a Likert 5-point scale, ranging from 1 (indicating very easy) to 5 (indicating highly difficult). The following table summarizes the discussed literature.

**Table 3 Text Difficulty Measurements in Translation and Interpreting Studies**

| Scholars                   | Measures              | Indicators                                                                                                                                           |
|----------------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dam (2001)                 | objective, subjective | specialised terms, numbers, average sentence length, clause length, words per clause, average rate of speech                                         |
| Jensen (2009)              | objective             | readability, word frequency, non-literal expressions                                                                                                 |
| Pavlović and Jensen (2009) | objective             | readability (Lix formula, SMOG scale)                                                                                                                |
| Liu and Chiu (2009)        | objective, subjective | information density, readability (Flesch Reading Ease), new concept density, experts' rating                                                         |
| Chang (2001)               | subjective            | experts' rating: translatability; bilinguals' rating: comprehensibility, readability                                                                 |
| Hvelplund (2011)           | objective, subjective | Readability (Flesch Reading Ease, ARI, Flesch-Kincaid, Coleman-Liau, Gunning Fog, SMOG), word frequency, non-literal expressions, evaluators' rating |
| Hild (2011)                | objective             | lexical parameters, syntactic measures, discourse parameters                                                                                         |
| Sjørup (2013)              | objective, subjective | Readability (Flesch Reading Ease), evaluator's rating                                                                                                |
| Wang (2017)                | objective             | readability, word frequency                                                                                                                          |
| Ma (2019)                  | objective, subjective | word frequency, experts' rating                                                                                                                      |

In the aforementioned literature addressing text difficulty measurement, it is discernible that scholars do not adhere to a singular standard for assessing text difficulty. However, readability and word frequency emerge as frequently referenced parameters. From the beginning of the 1920s, scholars studying reading have shown a propensity to create readability formulas<sup>4</sup> as

---

<sup>4</sup> Popular readability formula such as Flesch–Kincaid Readability test, Gunning Fog Index, Flesch Reading Ease Formula, SMOG Formula, Coleman-Liau index, Automated Readability Index etc.

tools for gauging the difficulty of texts, resulting in the publication of over 200 such formulas by the scholarly community as of the 1980s (Klare 1984). A readability formula is a statistical construct that amalgamates quantifiable features of a text, which optimally forecast the text's difficulty. This may include elements like the mean sentence length in terms of words or syllables, the average word length in characters, and the proportion of challenging words (Sun 2015). However, a weakness of using readability formulas is that they do not inherently establish causal links between selected predictor variables, such as vocabulary difficulty and sentence length, and the response variable – readability (Sun 2015). Moreover, the readability formulas discussed until now primarily cater to native English speakers, particularly those in America. This raises an essential query as to whether these predominant readability formulas are effective for trainees of English as a Second Language. Another disadvantage of readability, as argued by Jensen (2009, p. 68), is that “readability indices are only sensitive to a text’s level of complexity based on its surface structure”. The readability indexes, which rely on measurements of character or syllable length, have limitations in predicting the perceived difficulty of individual words or compound terms (Hvelplund 2011). Thus, relying solely on readability as the exclusive measure of text difficulty is frequently inadequate. In recent years, an increasing number of scholars conducting research in Chinese interpreting and translation have been experimenting with subjective evaluations to assess text difficulty. Word frequency is also considered as a strong indicator of the lexical levels of text difficulty. It is a “powerful measurement tool for estimating the relative amount of effort that needs to be invested in the cognitive processing of a particular word” (Jensen, 2009, p. 69). Marks et al. (1974) suggested that effective comprehension of a text largely hinges on the individual’s proficiency in the language’s vocabulary, with the semantic complexity of words constituting the primary determinant of difficulty disparities. Word difficulty is often associated with word frequency, and the frequency of a reader’s exposure to a specific word is deemed a crucial indicator of word knowledge (Ryder and Slater 1988), a concept commonly referred to as the word frequency effect. This effect, as articulated by Just and Carpenter (1980), is underpinned by a cognitive model suggesting that frequently utilized words maintain a higher level of baseline activity, hence necessitating minimal activation for their retrieval from an individual’s mental lexicon. High-frequency words tend to be perceived and comprehended more rapidly and efficiently than their low-frequency counterparts (Howes and Solomon 1951; Balota and Chumbley 1984; Jescheniak and Levelt, 1994; Rayner and Duffy 1986). Word frequency impacts both the legibility and acceptability of a text, suggesting that a text abundant in high-frequency words is generally easier to comprehend than one filled with rarely used words.

The text difficulty assessment method used in this study will be presented in the methodology section.

#### **2.4.4 Working Memory Capacity as A Covariate Variables**

A covariate, by its nature, is a type of control variable that could potentially influence the outcome of an experiment but is not the primary variable of interest. Given the significant role working memory capacity (WMC) plays within cognitive load study and considering the evidence of its impact on interpreting performance in existing literature, we felt it was important to account for WMC in our investigation.

Simultaneous interpreting and sight translation task requires interpreters to comprehend the information presented in the source language and store it temporarily while concurrently producing that information verbally into the target language (Gerver 1976; Macnamara and Conway 2016). By the very nature of their operation, SI and ST greatly demand an individual's cognitive resources. To better understand the phenomenon of SI and ST, some researchers have tried to identify the cognitive resource that are crucial to interpreting efficiency. Working memory has been held to be one such important resource (Conway et al. 2005). Gile (1995, p. 167) indicated that “working memory is necessarily part of the language-comprehension process and of the speech-production process”. Working memory is considered to be a limited-capacity system that can temporarily store and process information (Baddeley and Hitch 1974). Just and Just and Carpenter (1992) suggested that individual differences in working memory capacity account for performance differences in language comprehension. Engle and Kane (2004) also argued that individuals with higher working memory capacity can better uphold mission goals and resolve conflict. In the past three decades, most studies have focused on whether WMC is different between simultaneous interpreters and non-interpreter controls, and evidence for professional interpreters show greater WMC than various comparison groups has been found (Darò and Fabbro 1994; Padilla et al. 1995; Christoffels et al. 2006; Köpke and Nespoulous 2006; Tzou et al. 2011; Babcock and Vallesi 2017). However, research on the correlation between working memory capacity and interpreting performance is often inconclusive or inconsistent. In addition, to my knowledge, there are no empirical studies on the relationship between sight translation and working memory capacity.

To date, researchers have not yet reached a consensus about the role of WMC in interpreting performance may primarily be due to various methodologies and cross-linguistic differences. Christoffels et al. (2003) used digit span and reading span (both English and Dutch version) test to measure the Dutch trainee interpreter's WMC and found that digit span was correlated with the interpreter's English–Dutch SI performance, but the reading span was not. Zhang (2012) conducted an experimental study in which he asked trainee students to perform a reading span test (Chinese version) and an English–Chinese simultaneous interpreting task. He divided the students into low-level WMC and high-level WMC groups based on the reading span results and found that low-level WMC and SI quality show a significant correlation, while the correlation between high-level WMC and SI quality is not significant. Zhang (2012) also drew the conclusion that the main effect of WMC on SI quality is not significant under the slow delivery rate condition, while in the case of the high delivery rate condition, the main impact of WMC on SI performance is significant. Lin et al. (2018) also used the reading span test in their study and claimed that WMC was positively correlated with both English–Chinese SI fluency and Chinese – English SI fluency and WMC factor can predict SI fluency. Differently, Injoque et al. (2015) asked Spanish-speaking professional English interpreters to undertake four memory span tests: digit span, listening span, digit span with articulatory suppression and listening span with articulatory suppression. They discovered a significant correlation between SI performance and both measures of WMC with articulatory suppression but not between SI performance and the same measures without articulatory suppression. Furthermore, Song and Li (2020) conducted an experiment using the speaking span (both English and Chinese version) to measure the Chinese trainee interpreter's WMC. In their experiment, trainees were also asked to perform an English–Chinese SI task, and they found that the results of the English speaking span was correlated with the interpreter's SI fluency, but the results of the Chinese speaking span was not correlated with interpreter's SI fluency. According to these results, we can find that the researchers used different memory span measurements to test interpreters' WMC, and the results were very different.

To sum up, in light of the influence of WMC on interpreting performance and given that interpreting performance is the focal point of this study, we have included WMC as a covariate variable. The intention is to discern whether it can account for differences when discussing the results.

## **Chapter 3 Methodology**

This chapter introduces the mixed-methods approach that underpins this study, setting the groundwork for understanding the comprehensive research methodology. Subsequent sections elucidate the specific methods applied in Experiment One, followed by those implemented in Experiment Two, offering a layered exploration of the research methodologies.

### **3.1 Mixed-methods Approach for Assessing Cognitive Load**

The term methodology describes the comprehensive strategies employed throughout the research process, focusing on the rationale behind data collection, the types of data gathered, the locations and methods of data collection, and the techniques used for data analysis (Hale and Napier 2013). Translation and Interpreting Studies encompass a wide range of research paradigms, with linguistics, discourse analysis, social sciences, and psychology having the greatest influence (Gile 2003; Hale and Napier 2013). This diversity is due to the fact that interpreting studies often borrow and adapt methods from various fields to suit their purposes. As Pöchhacker (2011) suggested, the methodology for an interpreting research project can adhere strictly to a single specific paradigm or adopt a more flexible approach by combining various paradigms through a mixed-methods strategy. By combining various research methods, we can obtain a more comprehensive understanding of the subject matter, ultimately leading to more accurate and reliable findings. Additionally, mixed methods can help to identify potential biases, limitations, or inconsistencies in the data, as well as uncover new insights that may not have been apparent when using a single research method. Therefore, mixed methods research provides a more holistic understanding of the complex processes and practices in Translation and Interpreting Studies. This study consists of two experimental parts: an online experiment and a lab experiment. A mixed-methods approach is applied to both experiments.

### **3.2 A Within-subject Design**

A within-subject design has been adopted for both experiments. At its core, a within-subject design allows for a robust analysis of the impact of different conditions on the same set of participants. This contrasts with a between-subjects design, wherein each participant is exposed

to only one condition. One of the significant advantages of a within-subject design is that it does not require as many participants as a between-subjects design, and it can also help reduce the effect of individual differences (Cherry 2022). This reduction in variability is due to the fact that each participant serves as their own control, mitigating confounding variables such as age, gender, or educational background. However, this kind of design also has a drawback, that is, the participant's performance may improve or deteriorate from one condition to the other because they are completing the ST and SI tasks multiple times under different conditions during the experiment. These are called “practice effects (improvement) and fatigue effects (deterioration)” (Guo and Yang, 2019, p. 75). The former is because the participants might become more proficient in conducting tasks as they gradually get more familiar with the experimental environment and task, while the latter is because the participants’ performance might get worse because of the boredom or fatigue with the task. Considering these challenges, it becomes paramount to apply counterbalancing. Therefore, the order of the ST and SI tasks was counterbalanced across participants, with some starting with ST before SI, and others vice versa. Furthermore, within each ST and SI task, the sequence of the four texts was randomized to eliminate potential order effects. In this regard, a well-structured and meticulously executed within-subject design stands as a powerful tool in the experimental research, maximizing the integrity and validity of results.

### **3.3 Statistical Method for Analysing Experimental Results**

In both Experiments, I used statistical methods to analyse data. All statistical analysis of quantitative data were conducted in Jamovi software. Statistical methods play an important role in explaining experimental results. As “the immediate result of an experiment is a large series of numbers that represent behaviour under different conditions” (Kantowitz et al., 2008, p. 78), we need to use statistical analysis to understand which data are reliable (not by chance) and further test our hypothesis. We can refer to what psychologists said about the role of statistics in analysing experimental results:

... scientists believe that all data are tainted at birth. Data belong to Chance or to Science — but never to both. Before the psychologist can be sure that data belong to Science, the demon Chance must be exorcised. This is done by a ritual called inferential statistical analysis. ...Statistics are never a substitute for thought. Statistical analysis is a theoretically neutral procedure that serves theory and hypothesis testing. (Kantowitz et al., 2008, p. 78)

To test my hypotheses, I will report statistical analyses of the data, using Repeated Measures ANOVA, Linear Mixed Model and Correlation analyses. It is important to explain the logics behind these methods and how they contribute to my analysis when presenting the data results.

Firstly, a repeated measures ANOVA is the abbreviation of the Analysis of Variance and it is a “statistical procedure for testing variation among the means of more than two groups” (Aron et al., 2014, p. 351). The repeated measures ANOVA applies to experimental design with more than two conditions and variables. Before interpreting the findings, it is important to know what significant difference is and why we need to know if our analyses were statistically significant or not. During an experiment, the null hypothesis ( $H_0$ ) plays an essential role in significant testing, when we find significant effects, we can reject the null hypothesis and support our experimental hypothesis. The statistical significance shows “how confident are you that what you found in your small sample of participants reflects what is really occurring in the entire population” (Bourne, 2017, p. 99). When reporting an ANOVA, we use the  $p$  value to show the level of significance of an effect as expressed in the  $F$  ratio. The  $F$ -ratio is computed by dividing the variance and a large  $F$  ratio means that the difference between groups is much greater than you would expect. A  $p$ -value, also named probability value, is a measure that quantifies the probability of obtaining results at least as extreme as the ones observed, assuming that the null hypothesis is true. It does not directly indicate whether the observed difference is due to chance alone but rather provides a scale of evidence against the null hypothesis (Healey 2015). The guidelines in the relationship between  $p$ -value and significant difference is as follows:

- If  $p$  value  $> 0.10$  —→ not significant
- If  $p$  value  $\leq 0.10$  —→ marginally significant
- if  $p$  value  $\leq 0.05$  —→ significant
- if  $p$  value  $\leq 0.01$  —→ highly significant

When we find a significant effect in an ANOVA, we will use a method of contrast to know where the difference came from, to know what conditions are different and what is not different. The Post-Hoc comparisons can help us to answer these questions. Post-hoc comparisons also known as posteriori comparisons, because they are not planned in advance and come after the fact. It is a procedure conducted as part of an experimental analysis after an ANOVA. “In post-

hoc comparisons, all possible comparisons have to be taken into account when figuring the overall chance of any one of them turning out significant” (Aron et al. 2014 p. 377).

In the present study, I also employed a Linear Mixed Model to analyse the influence of various factors and ascertain the effect size. Linear Mixed Model is a form of regression analysis designed to handle data that have correlated observations, often arising due to hierarchically structured data or repeated measurements (Gelman and Hill 2007). When we are interested in analysing the influence of various factors and understanding the effect size, a Linear Mixed Model provide invaluable insight. Just like in linear regression, the Linear Mixed Model yield regression coefficients for fixed effects that tell us about the magnitude and direction of the relationship between predictors and the outcome variable. A Linear Mixed Model also allow for comparison between models with different fixed and random effects. This is useful to determine which factors (or interactions between factors) significantly contribute to explaining variability in the outcome (Bates and DebRoy 2004).

A correlation shows the association between scores on two variables. More precisely, “the usual measure of a correlation describes the relationship between two equal-interval variables” (Aron et al. 2014 p. 488). In this experiment, I tested whether or not a participant’s working memory capacity has an effect on their performance (accuracy scores). Therefore, I carry out the Pearson correlation, which is aiming to see “if there is a significant relationship between two of these variables, and to identify the strength of this relationship” (Bourne, 2017, p. 171). The correlation between two variables can result in one of three outcomes: No Correlation, Positive Correlation or Negative Correlation. A positive correlation means that greater working memory capacity is associated with greater accuracy scores and a negative correlation means that greater working memory capacity is associated with lower accuracy scores.

### **3.4 Methodology for Experiment One**

In the Experiment One, trainee interpreters’ cognitive load was evaluated through both subjective and performance methods. The subjective assessment involved participants gauging their perceived cognitive load, while the performance-based assessment focused on the accuracy score of the output in both SI and ST tasks. Additionally, the working memory

capacity of participants was determined using the Corsi test. These approaches ensure a comprehensive understanding of the trainees' cognitive load during the experiment.

#### **3.4.1 Subjective Method: The Self-Assessment of Perceived Cognitive Load**

Combining self-assessment with other methods, such as expert evaluations or retrospective protocols, can help researchers gain a more comprehensive understanding of the interpreting process and performance. Self-assessment allows participants to report their perceptions of their cognitive load, which can help me to understand the factors contributing to perceived cognitive demands and identify potential areas for improvement or further investigation. This method often used in Translation and Interpreting Studies (such as Gile 1995; Kiraly 2000; Tiselius 2009). It is important to note that self-assessment relies on subjective perceptions of cognitive load, so combining this data with other measures (e.g., performance measurement) will reduce the bias of the subjective assessment.

In Experiment One, cognitive load was assessed using a modified version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX). While the original NASA-TLX employs five 7-point scales to gauge task load level, it results in twenty-one gradations based on high, medium, and low estimates for each point (NASA ARC 1980). In contrast, the version adapted for Experiment One utilized a more streamlined 10-point scale. This revised scale ranged from 1 (very low) to 10 (very high). The original version of the NASA-TLX can be found in the Appendix. The decision to adapt the original NASA-TLX scale to a more straightforward 10-point format was driven by multiple considerations. Firstly, participants were required to complete the self-assessment after each condition, leading them to interact with the scale six times throughout the experiment. A 10-point scale is not only easier for participants to quickly comprehend and respond to but also expedites the response process. Moreover, considering the central theme of the study, it was deemed essential to minimize the cognitive load on participants. The original NASA-TLX twenty-one gradations demand more cognitive effort from participants to discern between nuanced choices, whereas a basic 10-point scale may streamline this decision-making process.

#### **3.4.2 Performance Method: Accuracy Evaluation**

This section will focus on how to construct an accuracy evaluation framework for this study. The performance-based techniques for measuring cognitive load include interpreting and

translation quality evaluation. The quality of interpreting and translation is a multifaceted and complex topic that scholars have approached from different perspectives, but a universal consensus on its definition and criteria for evaluation has not yet been reached (Kurz 1993; Gile 1995; Shlesinger 1997; Wadensjö 1998; Zwischenberger 2010; Pöchhacker 2001, 2016). Assessment of interpreting and translation are characterized by a lack of systematic methods and consistent standards, which results in a failure to establish objective scoring criteria (Sawyer 2004). Liu (2013) also argued that, although quality should be defined by observable and distinguishable constructs for measurability, these constructs often remain vague and are not operationalized in actual English-Chinese interpreting and translation assessment practice. Despite these assertions, the variability and subjectivity in evaluation do not arise solely from the flaws in assessment methods. In fact, the role of the evaluators and the purpose of the assessment play pivotal roles in influencing the outcomes. As some scholars (Garzone 2003; Ais and Becerra 2015; Pöchhacker 2016) suggested, the difficulty in defining interpreting quality stems from the need to consider disparate aspects involving various actors, who have different aims and perceptions of quality, thereby necessitating the use of diverse evaluation approaches. While disparities in assessment exist, they underscore the significance of establishing more context-specific and purpose-driven criteria in quality evaluations.

The present research encompasses a comparative analysis of two modes: simultaneous interpreting and sight translation. Despite the differences in their input methods, both modes produce oral outputs. This distinction necessitates an emphasis on evaluation criteria specific to oral interpreting rather than those created for written translation. Furthermore, given that the primary objective of this study is to understand cognitive load, as evidenced by performance measures, the quality assessment of SI and ST should consider criteria relevant to cognitive studies. This approach ensures that our evaluations are not only align closely with the inherent cognitive demands but also provide a comprehensive perspective on the variations in trainees' cognitive load during SI and ST tasks.

In the realm of cognitive studies related to interpreting, as mentioned in the Literature Review Chapter, many scholars (Gile 2003; Liu 2009; Seeber 2013; Korpál 2016) frequently emphasize the concept of accuracy as a key metric. Typically, in these studies, the quality of interpreting often gauged through accuracy, serves as a pivotal measure to understand cognitive load from a performance perspective. However, when employing accuracy as an evaluation criterion, what elements should be taken into consideration for its definition? Surprisingly, despite being

utilized as an evaluation criterion, accuracy lacks a uniform definition concerning its constituents and the methods employed for its measurement (Féal 1990; Gile 2003; Tiselius 2015; Pöchhacker 2016). When assessing an interpreter's output, the primary focus is the relationship between the source text and its rendition in the target language. Two dominant perspectives in this context are product-orientation and interaction-orientation, both of which considerably influence quality standards and assessment criteria (Pöchhacker 2016). Gile (1991) and Jones (1998) advocated that the concept of accuracy is associated with the product-oriented perspective because it focuses on the interpretation or target-text as a faithful image or exact reproduction of the original discourse. Gile (2009) further expanded on this by suggesting that accuracy is a combination of 'content' (information transfer) and 'packaging' with varying weights. However, according to Tiselius (2015), how much information is enough and what constitutes understandability with respect to accuracy in a specific situation still requires further examination. Such unresolved intricacies explain why very few scholars "have investigated accuracy as a construct in its own right or ventured evidence-based definitions" (Tiselius, 2015, p. 4). While the complexities surrounding the concept of accuracy persist, this study does not aim to offer a conclusive definition for accuracy. As highlighted earlier, due to the lack of a uniform standard for accuracy, it becomes imperative to shape evaluation criteria based on specific research purposes of the present study.

When discussing the measurement of accuracy, some scholars have quantified interpreting output. Barik (1973) utilized an error system encompassing omissions, additions, and substitutions to gauge accuracy. Mackintosh (1983) evaluated accuracy based on the semantic equivalence of meaning units. However, many have criticized the use of error counts due to the blurred boundaries in error categorization and the lack of precise definitions; more crucially, evaluating by units can be overly subjective and might label strategic omissions or additions as mistakes (Falbo 2015; Pöchhacker 2016). In fact, the omissions and additions that do not interfere with the delivery of crucial information from source text and are used to meet the challenges of task complexity should not be treated as a mistake. I prefer to label such omissions as 'positive omissions and additions'. Some scholars (Viaggio 2002; Visson 2005; Pym 2009; Korpala 2012; Barghout et al. 2015) argued that certain types of omissions might be a conscious decision made by the interpreter to confronted with extreme working conditions (e.g., fast delivery rate task). Nolan (2005) and Jones (2008) also both suggested that in extreme speech conditions, the interpreter should simplify the sentence and omit some "elegant

flourishes” (Jones, 2008, p. 25) when necessary. ‘Positive omissions and additions’ might also help interpreters release their working memory burden.

Altman (1994), expanding on error analysis, proposed that the severity of an error is determined by “the extent to which they affect the communicative impact of the speaker’s message” (1994, p. 30). Altman emphasized the communication process, underscoring the significance of considering contextual and situational elements when pinpointing interpreter errors. Falbo (2015) highlighted this is a notable advancement in error analysis, since it allows researchers to differentiate between omissions leading to a complete or partial loss of ideas and those that simply eliminate redundancies. Delivering source message’s meaning is frequently mentioned in the context of interpreting accuracy. Sergio (1991) also advised trainees that focusing on managing units of meaning, as opposed to individual words, can boost the efficiency of their short-term semantic memory. This insight is particularly relevant for our study, as our participants are trainees. In studies by Bühler (1986) and Kurz (1989) on interpreting quality, “Sense consistency with the original message” ranked first in both studies, while factors like “native accent” and “pleasant voice” (Kahane, 2015, p. 2) ranked last. In Liu’s (2001) research on interpreting expertise from the perspective of working memory analysis, she even asked the raters not to be influenced by the interpreter’s voice quality and accent when rating interpreter’s accuracy. In addition to this, when Seeber (2007, p. 293) was discussing interpreting accuracy, he advocated that “semantic aspects must be taken into account in order to offset the constraints imposed by a purely literal approach.” This way is use for balancing the restrictions of a purely literal approach.

In summary, when evaluating accuracy, my primary focus is on the trainees’ ability to convey the overall meaning of the source information while maintaining semantic coherence. The two markers assessed the extent of omissions and additions, examining whether they qualify as effective strategies. This decision is in line with the study’s emphasis on cognitive load. Specifically, we choose not to classify strategic omissions and additions as errors because we view them as the choices trainees make to manage their cognitive resources more efficiently. In this study, evaluating interpreting accuracy is, in part, assessing their strategic allocation and management of cognitive resources under a given task. If a student achieves a high score in interpreting accuracy, it indicates they have effectively managed their cognitive resources and made strategic decisions. This not only results in improved performance but also in a decreased

cognitive load. This observation supports the hypothesis that better performance is associated with a reduced cognitive load.

### **3.4.3 Working Memory Capacity Measurement: Corsi Span Test**

The memory span test chosen in the online experiment is the Corsi span test, which was developed by Philip Michael Corsi in the 1972 as a measure of visuospatial working memory capacity. It is very similar to the digit span task and defined as the longest sequence that a person can correctly repeat (Kessels et al. 2000). The Corsi Span Test complements other memory span tests by targeting a distinct aspect of working memory: visual-spatial processing. The Corsi Span Test involves both processing and storage components, as participants must encode the spatial-temporal positions of the blocks and maintain them in working memory while reproducing the sequences (Jones et al. 1995). In the Experiment One, I built 15 ‘repetitions’ for stimuli, in which the participants saw nine arranged blocks of different colours and needed to tap/click a sequence of blocks with a right colour (detailed settings can be found in the research design section). This dual-task nature of the test makes it particularly well-suited for assessing the dynamic interplay between processing and storage in visuospatial working memory (Berch et al. 1998).

The Corsi Span Test is sensitive to individual differences in visuospatial working memory capacity, enabling us to effectively assess and compare the abilities of different participants. This sensitivity makes the test valuable for investigating the cognitive processes underlying visuospatial working memory and for exploring the relationship between visuospatial working memory capacity and other cognitive abilities. Furthermore, Pham and Hasson (2014, p. 468) suggested that “The visual-spatial sketchpad is primarily used as storage not only for visual and/or spatial input but also for linguistic input that can be recoded into nonverbal or visual forms”. This means visual-spatial working memory plays a role in creating and maintaining mental images during language processing. When we read or listen to a description or sentence of a topic, we may form mental images to help us better understand and remember the information. Visuospatial working memory allows us to construct, manipulate, and maintain these mental images while we continue processing the linguistic input such as simultaneous interpreting. In addition, Swanson et al. (2001) argued that visuospatial working memory is correlated with reading comprehension. Research by Bayliss et al. (2003) and Goff et al. (2005) indicates a notable link between visuospatial working memory and reading comprehension.

Their studies using the Corsi task found that visuospatial working memory contributes independently to reading comprehension skills. Sight translation, an integral part of this study, inherently involves reading comprehension. This process demands the ability to monitor text layout, maintain the position of the current word, and systematically shift attention from one line to the next. Such tasks heavily rely on visuospatial working memory. Given the importance of these processes in sight translation, we chose the Corsi Span Test to evaluate the working memory capacity of sight translators.

In summary, the Experiment One employed self-assessment as the subjective method to measure trainee interpreters perceived cognitive load. The holistic accuracy evaluation of trainee interpreters' output served as the performance-based measure of their cognitive load. Additionally, the Corsi test was utilized to assess the trainees' WMC.

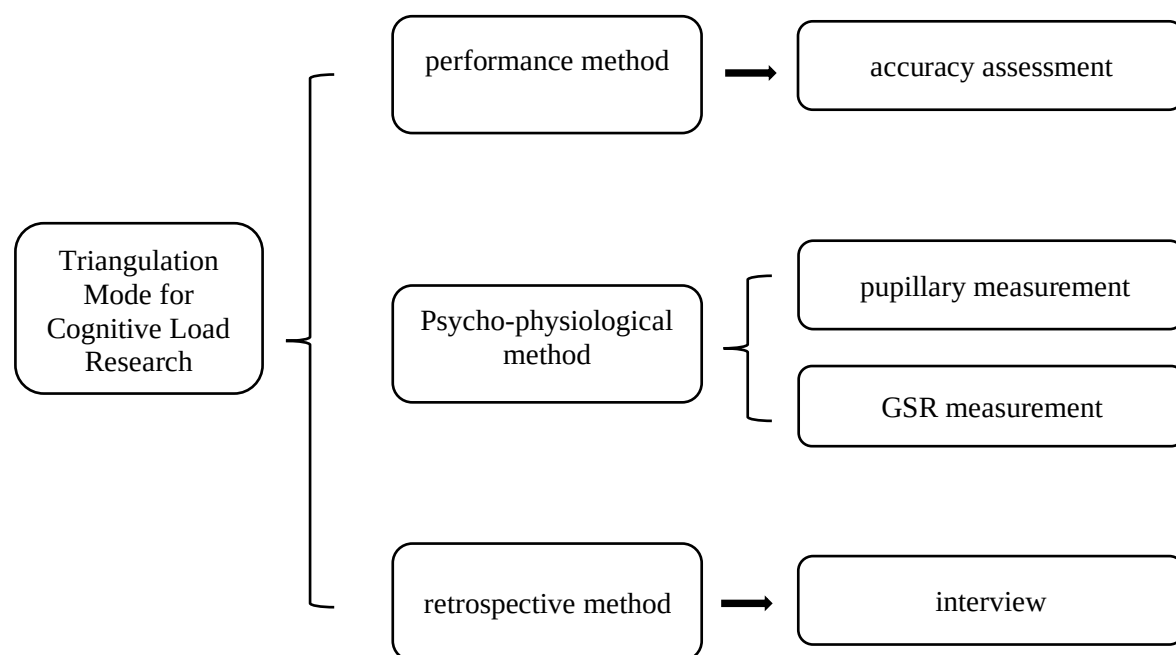
### **3.5 Methodology for Experiment Two**

The Experiment Two introduced several refinements in the data collection method, the first of which involves integrating both quantitative and qualitative methods to measure trainees' cognitive load. From a quantitative perspective, the experiment collected numerical data on pupil size, GSR, and accuracy scores. Additionally, qualitative insights were obtained through participant interviews regarding their performance in the interpreting tasks. This approach enables us to delve into the intricacies of trainees' interpreting strategies, and decision-making processes, and how these may influence their behaviour. Such experimental research designs are also highly effective in examining the process, product, and expertise involved in interpreting, as well as in evaluating hypotheses derived from genuine qualitative data, such as recordings of interpreted interactions (Hale and Napier 2013).

Secondly, aside from the approach integrating both quantitative and qualitative data, we also employed the principle of triangulation. Triangulation, in the realm of Interpreting Studies, typically involves the convergence of various research methods to bolster the validity and comprehensiveness of findings. While the combination of quantitative and qualitative methods provides a broader spectrum of data, triangulation ensures that these data points corroborate and reinforce one another, enhancing the reliability of our conclusions. In Carl's (2009, p. 225) words, triangulation is "the application of combined research methodologies, theories or data

sources to double (or triple) check scientific results”. Every research method has its own set of limitations and biases, triangulation increases the reliability of research findings by cross-validating results from multiple methods or data sources and counteract some limitations. Furthermore, method bias or participant bias can occur in various forms in research. Triangulation helps to minimize the impact of these biases on research findings by incorporating multiple methods, data sources, or perspectives (Williams and Chesterman 2002; Carl 2009; Hale and Napier 2013).

In the Experiment Two, the triangulated framework for this study is presented below:



The triangulated approach adopted, anchored in the principles of Interpreting and Translation Studies, offers a multifaceted perspective to thoroughly analyse trainee interpreters’ cognitive load and strategies. By incorporating multiple methods, this methodology not only delves into the intricacies of their interpreting behaviours but also seeks to provide a comprehensive response to the research questions. In doing so, the credibility and validation of the study’s findings are enhanced.

### 3.5.1 Performance Method: Accuracy Evaluation

The evaluation criteria and methods for Accuracy Evaluation in Experiment Two are identical to those in Experiment One; thus, further elaboration is dispensed with for brevity.

### **3.5.2 Psycho-physiological Method: Pupillary Measurement**

Pupillometry equipment has evolved over time, with modern devices offering advanced features, higher precision, and better ease of use. Pupillometric techniques are used to measure changes in the pupillary aperture of the eye's diameter in response to psychophysical or psychological stimuli (Granholm and Steinhauer 2004). The advantage of analysing pupil behaviour and pupillary responses to specific stimuli is that pupillary measurements are non-contact, easily accessible, and objective, requiring only slight cooperation from the participant being examined (Kelbsch 2019). The primary focus of pupillometry equipment is to accurately measure and track changes in the size and behaviour of the pupil. The most commonly used devices for pupillometry today are video-based eye trackers. In this lab experiment, the equipment provided by Psychology Department at the University of Stirling is Tobii 1750. It is an eye-tracking device developed by Tobii Technology and utilizes near-infrared illumination to enhance the contrast between the pupil and the iris (Tobii Technology 2007). The device emits infrared light that is invisible to the human eye, ensuring that the participant remains comfortable and undisturbed during the tracking process. The infrared light allows the cameras to clearly detect the boundaries between the iris and the pupil, ensuring accurate measurement of the pupil diameter (Laeng et al., 2012). The dual-camera setup allows the device to track both eyes simultaneously, providing a more comprehensive understanding of the participant's gaze behaviour and pupillary response (Laeng et al., 2012). Furthermore, this device uses advanced eye-tracking algorithms to analyse captured images and extract information such as pupil size, position and behaviour, and these algorithms can track even small changes in pupil diameter (Tobii Technology 2007). These data can be used to compare changes in trainee interpreters' cognitive load under different conditions.

It should be noted that while pupillometry is a valuable non-invasive tool for studying cognitive processes, there are certain limitations that we must take into account when interpreting results and designing experiments. To ensure that this type of metric is valid and effective, we must be extremely careful when collecting participants' pupil data. A good preparation is very important, which is based on my experience in pupil data collection during my master study. Human pupils are very sensitive and susceptible to external environment and body conditions (Li 2019). Light conditions, sound, vibration, stress or some other personal factors such as eye

colour, eye make-up or whether subjects drank coffee before the experiment can confound the results (Goldberg and Wichansky 2003; Hvelplund 2014; Teixeira et al. 2014; Li 2019).

Drawing from my academic journey, in 2019, during my Master's study at the University of Stirling, I employed eye-tracking techniques for my dissertation titled "*The Effect of Syntactic Difficulty on Cognitive Effort in English–Chinese Sight Translation: An Eye Tracking Study*". In that study, I included pupil size as an indicator, envisioning it as a potential metric for cognitive effort. However, the outcomes were not in line with our anticipations. After post-experiment dialogues with the participating students and a review of the methods, I realized that in experiments involving pupil data collection, controlling for potential influencing factors is extremely essential. Even though the experimental device and settings from that study differ significantly from my current PhD research, it underscored the necessity of careful preparation. Therefore, to ensure valid pupil size measurements, all participants were told not to wear contact lenses or makeup and not to drink any caffeinated or sugary drinks before the experiment. Upon arrival, participants were given some time to settle into the lab environment before the experiment was started. As a further measure, I also engaged them in casual conversations during this period to facilitate a relaxed environment before commencing the experiment. The calibration process for eye tracking devices would be explained in detail in the research design section.

### **3.5.3 Psycho-physiological Method: GSR Measurement**

We employed a Galvanic Skin Response (GSR) system in Experiment Two. The Biopac MP36 was utilized to capture GSR data. This system provides both wired and wireless physiological measurement tools, allowing for comprehensive in-lab analyses of cognitive activities. The SS3L GSR transducer was employed to monitor skin reactions, which was affixed to the index and middle fingers of participants' hands. This GSR methodology offered a non-invasive and real-time metric, complementing other tools in gauging participants' cognitive load throughout the tasks.

### **3.5.4 Combining GSR and Pupil Dilation for Psycho-physiological Measurement**

In this section, I will explain the reason of combining GSR and pupil size measurement in the Experiment Two. Combining these two measures represents a multimodal assessment approach, which integrates multiple psycho-physiological indicators to better understand

human cognitive states. Multimodal assessment is an approach that involves the collection and analysis of data from multiple sources or modalities to provide a more comprehensive understanding of the physiological state or behaviour of an individual (Borsboom 2005). In the context of psychology or cognitive neuroscience, it typically refers to the simultaneous use of multiple methods or tools to measure different aspects of human cognition, emotion, or behaviour (Cacioppo et al. 2008; Cowley et al. 2016). Multimodal assessment is useful because it can provide a richer and more accurate picture of an individual's cognitive or emotional state. By combining data from different modalities, researchers can overcome the limitations of single-method approaches, which may suffer from biases or inaccuracies. Moreover, multimodal assessment can help to identify patterns and relationships that might be missed when only one method is used.

As mentioned above, GSR is associated with the autonomic nervous system (specifically the sympathetic nervous system) and the sweat glands activity, and differs from pupillometry, which is influenced by both the sympathetic and parasympathetic nervous systems. Although GSR and pupillometry are both physiological measures used to assess emotional, cognitive, and physiological states, they are based on different physiological mechanisms. In this context, the combination of GSR and pupil dilation can help reduce the impact of potential confounding factors or environmental influences, that may affect one measure more than the other. Therefore, a multimodal approach combining GSR, and pupil dilation can help overcome the limitations of individual measures and improve the overall accuracy and reliability of cognitive load assessment.

### **3.5.6 Retrospective Method: Interview**

Interview was used as a retrospective protocol in the Experiment Two. Introspection, also known as subjective observation at times, is among the initial approaches utilized in the study of psychology. It remains the sole method that enables individuals to directly examine their cognitive processes as they occur – through the real-time verbalization of these processes (Seeber 2015). As SI and ST involves processing both input and output at the same time, it is not feasible to add another auditory output stream (e.g., think aloud protocol) while working in the simultaneous interpreting or sight translation modes. Therefore, SI and ST have primarily been investigated through the use of backward or delayed introspection, producing

retrospective protocols (Seeber 2015). Engaging in this form of retrospection is equivalent to recollecting and examining past experiences that have been committed to memory.

In the Experiment Two, a post-experiment interview was conducted as a form of retrospective protocol. This interview focused on the trainees' thought processes, strategies, and decision-making that occurred during the interpreting task, providing valuable insights into their cognitive processes and actions. To facilitate a more accurate and comprehensive recollection of the SI and ST process by the participants, the interview was conducted immediately following the experiment. A semi-structured interview was applied into the Experiment Two. It is a qualitative data collection method that combines the flexibility of an open-ended interview with the directionality and organization of a structured questionnaire. It typically involves using a predetermined set of open-ended questions, along with the freedom to diverge and explore other topics as they arise during the conversation (Bryman, 2016). The flexibility of a semi-structured interview allows trainees to share their interpreting experiences in depth, offering insights that might not be captured through strictly structured questions. In addition, while the predetermined questions ensure that core topics of interest are covered, the flexibility to follow up on unexpected responses ensures that no pertinent information is overlooked. Take an example, a predefined question in the interview is "Which factor do you believe has a greater impact on your in ST task: delivery rate or text difficulty? Why?" One participant responded, "I think it is delivery rate, largely due to my daily training routine." To probe further and derive more context, a follow-up question was posed: "Can you describe the routine of your daily ST training? Is it a routine you have crafted on your own, or do you follow by the guidance in your training course?" This example demonstrates the flexibility of a semi-structured interview, which starts with a broad question and then allows for more specific follow-up questions based on the participant's answers. Thus, interviews as a retrospective protocol can offer valuable insights into interpreters' cognitive processes and decision-making during their simultaneous interpreting and sight translation.

### **3.5.7 Working Memory Capacity Measurement: Word Span Test**

Instead of the Corsi Span, the Experiment Two utilized the Word Span Test to measure the participants' WMC. Refer to the Discussion Section of Experiment One for an explanation regarding this modification. Word span test is a psychological assessment tool designed to measure an individual's working memory capacity for verbal information (Rosen and Engle

1997). The word span task is particularly relevant to the phonological loop, which is responsible for the temporary storage and manipulation of verbal and auditory information. Similar to the Digit Span Test, the word span test involves presenting participants with a series of words instead of digits, and the participants are asked to recall the words in the same order they were presented (forward span). The purpose of word span task is to assess people's verbal working memory capacity and to study the role of language and semantic processing in working memory. This test is also employed in studies examining the relationships between working memory capacity, language proficiency, reading comprehension, and other cognitive abilities. SI and ST require significant linguistic processing, both in understanding the source language and producing the translation in the target language. The Word Span Test, which involves the recall of sequences of words, is directly related to these linguistic processes. Unlike visuospatial tests like the Corsi Span, the Word Span Test direct emphasis on linguistic processing, making its findings potentially more applicable to SI and ST studies.

For the purpose of this study, both visual and auditory versions of the word span test were employed in the Experiment Two. The motivation behind this decision stems from the inherent differences in the cognitive processing demands of ST and SI. While ST predominantly involves the processing of written text, making it akin to the visual version of the word span test, SI majorly revolves around the processing of spoken language, aligning it closer with the auditory version of the word span test. Given the distinct input modes in SI and ST, it was deemed essential to incorporate both versions of the word span test to ensure a comprehensive assessment of interpreters' working memory capacities.

Additionally, the words selected for the Word Span Test in Experiment Two are detailed below. A total of 208 lowercase words generated with the MRC Psycholinguistic Database<sup>5</sup> (Version 2.00) were used in the visual WMC test (104 words) and auditory WMC test (104 words). All the selected stimuli are nouns, with each word containing between four and seven letters. These words vary in syllable count, ranging from one to four syllables. The settings for the word selection are inspired by the guidelines established by Pointe and Engle (1990) in their research. In the present Word Span Test, the frequency of selected words ranging from 100 to 400. Here, the measure of frequency reference to the frequency of occurrence of a word given in the norms

---

<sup>5</sup> MRC Psycholinguistic Database: [https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa\\_mrc.htm](https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm)

of Kucera and Francis (1967), with a maximum frequency of 69971 and a minimum frequency of zero (MRC Psycholinguistic Database 1997).

To summarize, in Experiment Two, various methods were employed to gain a comprehensive understanding of the trainees' cognitive load. Pupillometry and GSR were utilized for psychophysiological measurements. The accuracy of the SI and ST output was assessed for performance measurements. The word span test was employed to measure trainees' WMC, and retrospective interviews were conducted to understand trainees' decision-making and strategies during the SI and ST processes.

## Chapter 4 Experiment One

This chapter provides a detailed overview of the online experiment, referred to as Experiment One. The narrative progresses by offering a comprehensive view of the research design, subsequently portraying participant profiles, experimental stimuli, and the methodology for data treatment. The results segment showcases findings related to accuracy, WMC, and perceived cognitive load. A subsequent discussion juxtaposes these findings with prevailing literature, culminating in a reflective overview of the inherent limitations of the online experiment.

### 4.1 Research Questions

**Question 1:** Do text difficulty and delivery rate influence trainees' cognitive load in SI tasks?

**Question 2:** Do text difficulty and delivery rate influence trainees' cognitive load in ST tasks?

**Question 3:** Do trainees' cognitive load differ between ST and SI tasks?

**Question 4:** Does working memory capacity, when considered as a covariate variable, correlate with the cognitive load (accuracy result) of interpreting trainees in SI and ST tasks?

### 4.2 Independent, Dependent and Covariate Variables

- Independent Variables:
  - ◇ Task with the levels: sight translation (ST) and simultaneous interpreting (SI)
  - ◇ Delivery rate with the levels: normal vs. fast
  - ◇ Text difficulty with the levels: normal vs. difficult
- Dependent Variables:
  - ◇ ST and SI performance accuracy score
  - ◇ Perceived cognitive load level self-assessment score
- Covariate Variables: Working Memory Capacity (WMC)

### 4.3 Hypothesis in Experiment One

H1: Both text difficulty and delivery rate influence trainees' cognitive load in SI tasks.

H2: Both text difficulty and delivery rate influence trainees' cognitive load in ST tasks.

H3: Trainees experienced higher cognitive load in SI than in ST task.

H4: Trainees' working memory capacity as a covariate variable correlated with their accuracy score.

## **4.4 Research Design**

Experiment One was conducted online, and a within-subject design was applied to this online study. The decision to begin with an online format stemmed primarily from the restrictions imposed by the pandemic. This approach not only provided a viable solution in a constrained environment but also granted indispensable insights that significantly informed the subsequent lab experiment. The following part is about the experimental design.

### **4.4.1 Experiment Instruments**

#### **4.4.1.1 Online Experiment Platform**

In Experiment One, participants were required to equip with a computer/laptop, a mouse, and a headphone with microphone. The experiment was conducted using the Testable platform (<https://www.testable.org/>), a behavioural experiment tool developed by a group of academics and scientists, which notably offers support for Chinese. The psychology department at the University of Stirling bought a license, allowing their staff and students to utilize this platform for research purposes. Testable synchronously records participants' performance during a test, without the need for additional recording equipment. That is a great help when interpreting research fieldwork is not feasible during the pandemic. The eligible participants received a link by email to an online test and a tutorial from the researcher. A detailed tutorial (see Appendix 3) about how to conduct the test was very useful in increasing the effectiveness and availability of the data pool.

#### **4.4.1.2 Self-assessment of Perceived Cognitive Load**

In Experiment One, the questions in the self-assessment reference the contents of NASA Task Load Index (TLX). Participants were asked to complete a self-assessment of their perceived cognitive load levels right after completing the task of each condition. The purpose of doing this was to allow participants to immediately recall the level of perceived cognitive load they experienced during the task they had just finished. Thus, they had to complete six texts/videos

as well as six self-assessments, which allow us to compare participants perceived cognitive load level in different conditions.

#### **4.4.1.3 Corsi Span Test**

The WMC test was conducted by using PsyToolkit (<https://www.pytoolkit.org/>), which is a free-to-use toolkit for programming and running cognitive-psychological experiments.

#### **4.4.2 Participants**

Before the online experiment, information about participant recruitment was posted on my personal social networking sites (WeChat Moments and Instagram). Participants interested in the experiment will be provided with a preliminary questionnaire to evaluate their background and ascertain their suitability for the study. As a supplement to the recruitment information, the pre-participation questionnaire aimed to gather basic information on participants' academic backgrounds and training experiences. This included details about the university they were currently enrolled in, as well as the duration they had studied English before starting their postgraduate translation programme. Furthermore, the questionnaire inquired about any formal training they might have received in sight translation or simultaneous interpreting theories and practices. After the selection, a total of thirty-two Chinese trainees took part in Experiment One. Ten of the participants were paid, and twenty-two of them volunteered because they were interested in this project, especially the experimental methods employed. Their ages ranged from twenty-two to twenty-five years old, and three were male and twenty-nine were female. As most students enrolled in Master of Translation and Interpreting programmes are predominantly female, no attempt has been made to equalise the number of male and female participants. Although the number of female participants was higher than that of male participants, the similarities in their personal backgrounds and training experiences could minimise the potential effect of the gender imbalance. A profile of the participants is provided below.

**Table 4 Profile of Participants**

|                                  |                                                                |
|----------------------------------|----------------------------------------------------------------|
| Total Participants               | 32 trainees                                                    |
| Age Range                        | 22—25                                                          |
| Current Degree Programme         | MA/MSc Translation and Interpreting                            |
| Undergraduate Degree Programme   | Translation and Interpreting                                   |
| Native Language                  | Mandarin Chinese                                               |
| Secondary Language               | English                                                        |
| Age of Starting English Learning | 10-12 years                                                    |
| Training in SI & ST              | Received at least 10 weeks of relevant master course training. |
| Professional Experience          | None in interpreting or translation at the time of experiment  |

#### **4.4.3 Experimental Stimuli**

##### **4.4.3.1 Brief on Source Texts**

A total of six Chinese texts were used in Experiment One. All of them were excerpts from video transcripts, articles and websites. The order of the source texts was counterbalanced across participants. The six source texts were used in both the SI and ST tasks, but participants received different texts for each task. This means the ST texts done by some participants also served as the SI (video) done by other participants, while the SI texts (video) done by some participants also served as the ST texts done by other participants. There are six source texts used in Experiment One, with respective word counts of 122 words, 124 words, 123 words, 123 words, 123 words, and 122 words. Text A, B and C discussed eating habits and Text D, E and F discussed listening habits. The researcher changed the words, phrases and the word order of some sentences. The six texts were set as follows:

- **Two texts of Normal Difficult Text with Normal Delivery Rate**
- **Two texts of Difficult Text with Normal Delivery Rate**
- **Two texts of Normal Difficulty Text with Fast Delivery Rate**

#### 4.4.3.2 Evaluation of Source Text Difficulty

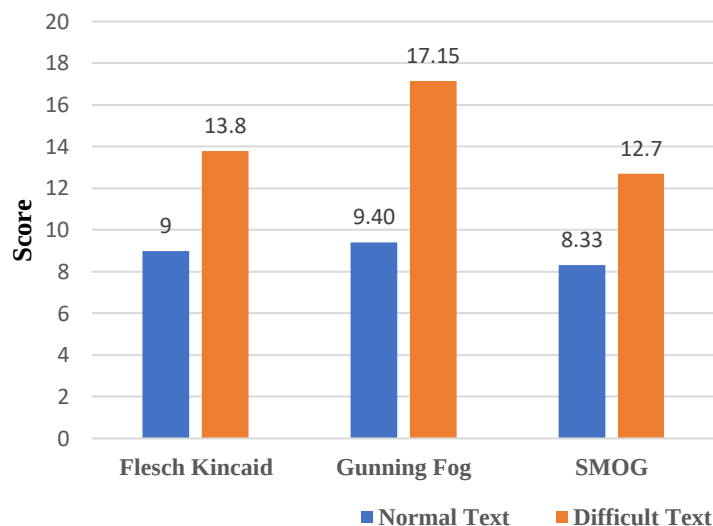
Experiment One utilized objective/quantitative methods to determine the difficulty of the source texts. Readability, a key measure, was determined through four widely used metrics: Flesch Reading Ease, Flesch Kincaid Grade Level, SMOG Index, and Gunning Fog Score. These metrics are standard tools in assessing text readability. Additionally, the Word Frequency Index served as another standard for quantitatively evaluating text difficulties.

#### 4.4.3.3 Readability Measurement

The Flesch Reading Ease, Flesch Kincaid Grade Level, and SMOG Index are the most used metrics in translation and interpreting research, specifically for controlling text difficulty. On the other hand, the Gunning Fog Score<sup>6</sup> is designed to estimate the number of formal education years a person requires to comprehend the text upon first reading. The average scores of four Normal texts and two Difficult texts are presented as follows:

**Figure 1 Readability Test Accuracy Scores**

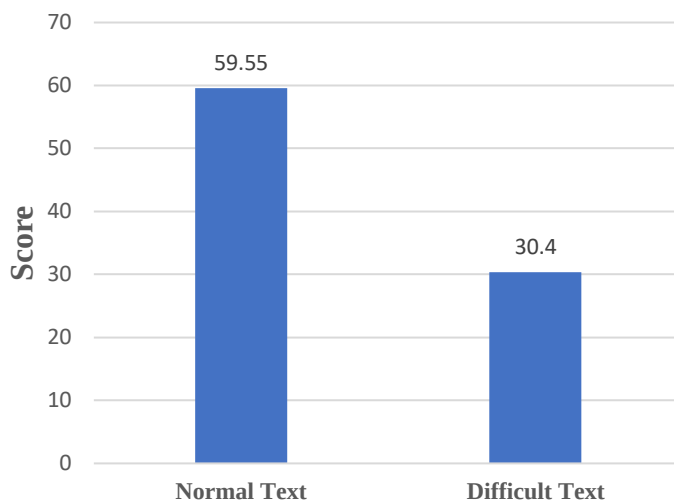
*Average Scores of Flesch Kincaid Grade Level, Gunning Fog and SMOG Index of Four Normal Texts and Two Difficult Texts*



<sup>6</sup> Flesch Reading Ease, Flesch Kincaid Grade Level, SMOG Index and Gunning Fog Score were measured by WebFX (<https://www.webfx.com/tools/read-able/>). This tool comes recommended by the UK Home Office for readability guidance, as seen here: <https://design.homeoffice.gov.uk/accessibility/readability>.

**Figure 2 Readability Test Accuracy Scores**

*Average Score of Flesch Reading Ease of Four Normal Texts and Four Difficult Texts*



In Experiment One, the application of the Flesch Kincaid, Gunning Fog, and SMOG indexes for readability assessment revealed a distinct escalation in complexity levels from Normal to Difficult text, as illustrated by the figures. Furthermore, the Flesch Reading Ease index assigns numerical scores ranging from 0 to 100. Higher scores are indicative of a text's ease of understanding, while lower scores suggest that a text is harder to comprehend. From the data presented, it can be observed that the average score for the Normal texts was 59.55, and for the Difficult texts, it was 30.4. According to Flesch (1948), scores ranging from 30 to 50 are categorized as 'difficult'.

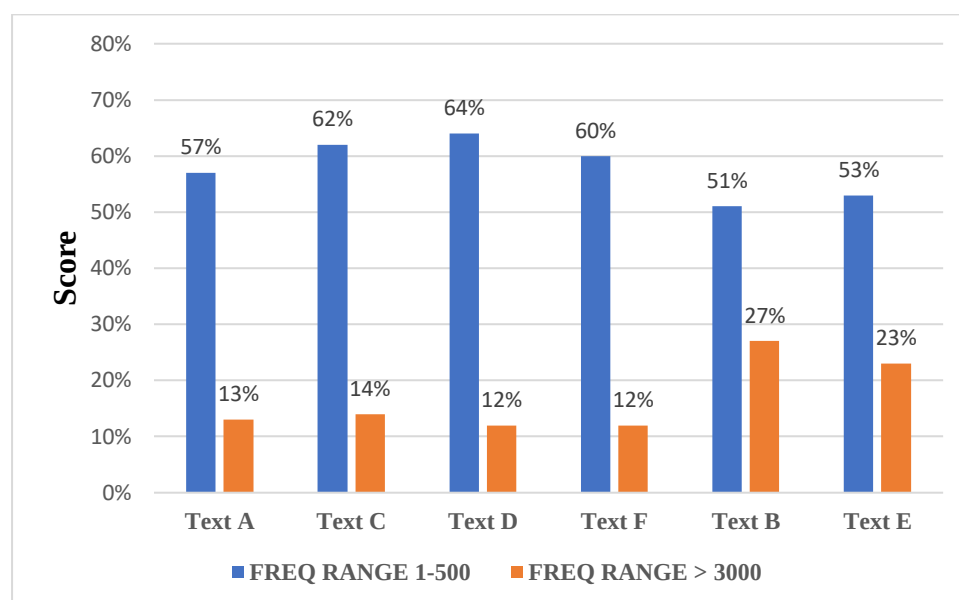
#### **4.4.3.4 Word Frequency Measurement**

The Word Frequency measurement was also incorporated into the present study. Drawing on the approach of Wang (2017), the tool (<http://www.wordandphrase.info/frequencyList.asp>) used to calculate the word frequency of each text in this study is identical to the one utilized in Wang's PhD thesis, which explored linguistic metaphors in English-Chinese translation. Here, words from the source texts were categorized into high-frequency and low-frequency groups. The criteria for these classifications are as follows: High-frequency words rank within the 1-500 range, whereas low-frequency words rank above 3000. In the subsequent figure, a comparison is made between the count of low-frequency words and that of high-frequency

words within both the Normal texts (Text A, Text C, Text D, Text F) and Difficult texts (Text B and Text E):

**Figure 3 Word Frequency Score**

*Word frequency scores of Each Text*



#### 4.4.3.5 Constraints on Delivery Rate and Presentation of Source Texts

In Experiment One, the delivery rate (including pause) settings were as follows:

- High delivery rate: 125 words per minute
- Normal delivery rate: 105 words per minute

The time pressure is presented in different ways in SI and ST. In SI task, the manipulation of time pressure can be reflected by speeding up the speaker's talking speed, but we cannot use this approach in ST task. To deal with this issue, in ST task, we control the variable of time pressure by constraining the time the sentence was presented on screen. In the sight translation (ST) task, each sentence remained on the screen for a duration equivalent to the time taken by the speaker to speak that particular sentence. To determine this duration, I recorded the time the speaker took to say each sentence. This approach ensures a more objective comparison of the trainees' output accuracy between sight translation and simultaneous interpreting. Additionally, this method mirrors some real-life ST scenarios where interpreters may have a constrained timeframe to process each sentence. Additionally, each sentence in the sight

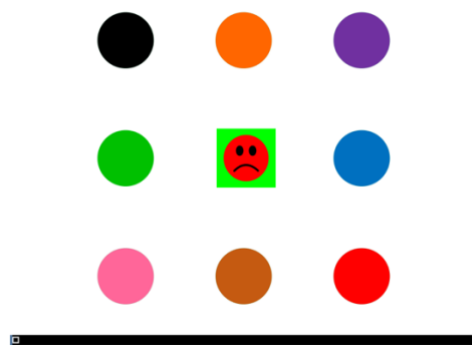
translation task was individually presented in consecutive slides. The reason for presenting one sentence in one slide is that the visual interference from the source text may retard comprehension and influence the quality of interpreter's performance (Song 2010). We hope to use this way to reduce the visual interference of long text to trainees and further reduce their visual pressure.

#### **4.4.3.6 Speaker in the SI Task**

This person is a native English speaker who has always resided in English-speaking countries. Additionally, the speaker has a basic proficiency in oral Chinese communication, having lived in China between 2011 and 2012. To ensure the most appropriate recordings were used for the study, the speaker filmed the speeches in English, creating three different video versions for each source text. These recordings were then sent to me for final selection.

#### **4.4.3.7 Stimuli for Corsi Task**

During the Corsi Span task, 15 sequences of colourful blocks appeared at different locations on screen and the participant was asked to remember the position and order of the sequence. The number of blocks in each group gradually increased from few to many. Once the sequence was shown, the participant would hear the word “go”, and then they needed to click the blocks in exactly the same order as shown previously. Then, when they finished, they needed to click the green “Done” button. They received immediate feedback, with a smiley face or frowny face indicating that they were correct or incorrect, respectively (See Figure 4).



**Figure 4 Frowny Face Means the Answer is Incorrect**

#### **4.4.4 Experiment Procedures**

As this is an online test, participants can join the test by clicking a one-time link. All they needed was internet, a computer and a pair of headphones with microphone, and the researcher

did not need to consider the participant's location when they conduct the online test. Participants were informed that their voices would be recorded throughout the duration of the experiment. The detailed procedures were as follows:

**Phrase 1: Experiment Guidelines and Consent Acknowledgement**

Participants were asked to read the tutorial to familiarise themselves with the online experiment's format and requirements. They then had to fill in and submit the consent form before the test began.

**Phrase 2: Sight Translation, Simultaneous Interpreting and Self-assessment Task**

Participants were prompted to sight translate and simultaneously interpret six source materials from English to Chinese. The order in which they engaged in ST and SI tasks was counterbalanced: some participants commenced with ST task followed by SI task, while others started with SI task and then moved to ST task. Furthermore, within each ST and SI task, the sequence of the three texts was randomized to eliminate potential order effects. After sight translating or simultaneous interpreting each text or video, they advanced to the next segment by clicking the 'NEXT' button. Intrinsically woven into both the ST and SI tasks was the self-assessment component. After finishing each condition - be it ST or SI - participants were prompted to gauge their perceived cognitive load. Over the span of both tasks, each participant undertook six such self-assessments, correlating to the three conditions of each type. After completing the ST and SI tasks, their performance was saved and uploaded automatically by Testable.

**Phrase 3: Corsi Span Test**

Following the completion of the SI and ST tasks, participants were presented with a link on the ending page. Upon clicking this link, they were directed to the final segment of the experiment – the Corsi Span Test. During this task, 15 sequences of colourful blocks appeared at different locations on screen and the participant was asked to remember the position and order of the sequence. The number of blocks in each group incrementally increased from a smaller to a larger count. Once the sequence was shown, the participant would hear the word 'go', and then they needed to click the blocks in exactly the same order as shown previously. Then, when they finished, they needed to click the green 'Done' button. After the working memory experiment, their data was automatically uploaded to PsyToolkit.

#### **4.4.5 Data Treatment**

##### **4.4.5.1 SI and ST Accuracy Scoring**

The holistic accuracy assessment was employed in this study. In the current study, the accuracy scores for participants' SI and ST were marked by two experts in Translation and Interpreting Studies. These experts, boasting nearly 15 years of experience in the field, are native Mandarin Chinese speakers and also proficient in English as their second language.

The agreement details between expert markers concerning the accuracy ratings are outlined as follows: Expert markers were assigned the task of independently evaluating the accuracy of participants' outputs. Before initiating the assessments, a pre-evaluation meeting was held to explain the marking criteria (refer to Appendix 5) and establish scoring expectations. This session also served to resolve any uncertainties related to the scoring process, thus ensuring a consistent understanding of the evaluation framework. A key agreement reached emphasised assessing trainees' ability to convey the overall meaning of the source material while ensuring semantic coherence. It was consensually agreed that strategic omissions and additions by the trainees, intended to enhance the coherence or relevance of the translated content, would not result in penalty points. Additionally, inspired by the insights from Liu (2001), both markers were informed that the comprehensibility of the translated output should be a consideration during evaluation. Hence, a SI or ST output deemed incomprehensible, even if it includes fragments of information from the source texts, would incur a deduction in score. Furthermore, to prevent evaluator fatigue—which could potentially affect the scoring's reliability—a recommendation was made for markers to take suitable breaks after reviewing five consecutive outputs. This strategy aims to ensure a consistent focus on detail and maintain accuracy throughout the evaluation process.

Inter-rater reliability was examined through a pairwise correlation analysis of two expert raters' evaluations of participants' accuracy in SI and ST. The analysis yielded results of  $r(30) = 0.987$ ,  $p < .001$ . The high correlation coefficient indicates an exceptionally strong agreement between the two raters in their assessments of SI and ST accuracy, underscoring the reliability of their marking.

#### 4.4.5.2 Self-assessment Data

For the self-assessment scores in Experiment One, each participant's total score was derived by summing the points from each question and then calculating the average. As mentioned in the methodology chapter, this method was based on our adapted 10-point scale for the NASA-TLX, which ranged from 1 (very low) to 10 (very high). This stands in contrast to the original NASA-TLX's more intricate twenty-one gradations based on its 7-point scale.

#### 4.4.5.3 Corsi Task Scoring

The scoring method according to the Span Length Measure. We measured the longest coloured dots/block's locations and sequences that participants could accurately recall. For example, if a subject can recall the correct location of five coloured blocks in order, the participant's span score is 5.

### 4.5 Result of Experiment One

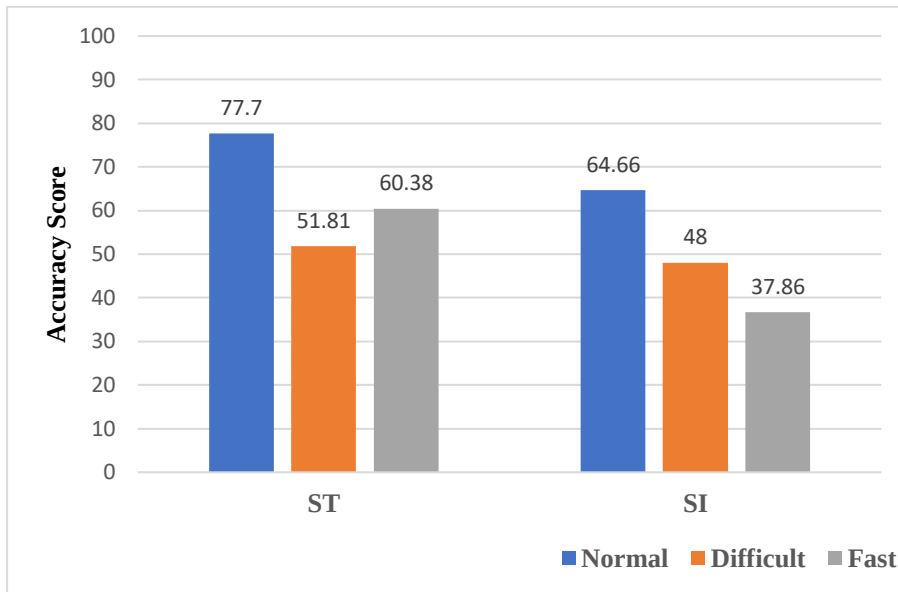
Before delving into the specifics, it is important to provide an overarching view of the outcomes from Experiment One. The results gathered offer pivotal insights into the core objectives of our study, shedding light on intricate patterns and tendencies. From now on, in order to make the description of the variables can be read clearly, I use '**Normal**' to represent 'normal difficulty text, normal delivery' and use '**Difficult**' to represent 'difficult text, normal delivery', use '**Fast**' to represent 'normal difficulty text, fast delivery', which is also how they will be shown in this chapter. All details of the whole results in Experiment One can be found in Appendix.

#### 4.5.1 Performance Analysis: SI and ST Accuracy Results

Shifting our focus to the details, an overview of the overall performance patterns observed in Experiment One shown in Figure 5. Specifically, for ST, the mean accuracy rates in the three different conditions were as follows: Normal (Mean = 77.70; SD = 9.28), Difficult (Mean = 51.81; SD = 12.81), Fast (Mean = 60.38, SD = 14.65). For SI, the outcomes were: Normal (Mean = 64.66, SD = 13.70), Difficult (Mean = 48; SD = 12.69), Fast (Mean = 37.86; SD = 8.03).

**Figure 5 Mean Score of ST and SI Accuracy**

*Mean Score of ST and SI Accuracy*



The ANOVA on translation scores in different conditions showed that the effect of task was significant ( $F(1, 31) = 66.2, p < .001$ ), showing that participants scored higher in the ST task (Mean = 63.30, SD = 16.40) than in the SI task (Mean = 50.17, SD = 16.07), and we can draw the findings that performance in ST was better than in SI (Mean Difference = 13.13). The interaction between task and conditions was also significant ( $F(2, 31) = 15.4, p < .001$ ), this result tells us that the differences between conditions (Normal, Difficult and Fast) in SI and ST task are not the same for each task. The effect of condition was significant ( $F(2, 31) = 116.6, p < .001$ ), and the further Post-hoc comparisons between the 3 conditions showed that in the ST task, accuracy score in the Normal condition (Mean = 77.70; SD = 9.28) was higher than in the Difficult condition (Mean = 51.81; SD = 12.81) and Fast condition (Mean = 60.38, SD = 14.65), the mean difference were 25.89 and 17.32, while the accuracy score in the Difficult condition was lower than in the Fast condition and the mean difference was -8.57 (see Table 5). The Post-hoc comparisons test in SI task also showed that accuracy scores in the Normal condition (Mean = 64.66, SD = 13.70) ranked highest among the three conditions, and the mean difference were 16.66 and 26.80. The accuracy score in the Difficult condition (Mean = 48; SD = 12.69) was higher than in the Fast condition (Mean = 37.86; SD = 8.03) and the mean difference was 10.14 (see Table 6).

**Table 5 Condition Pair Comparison in ST***Condition Pair Comparison in ST*

| Condition Pair     | Mean Difference | <i>P</i> value |
|--------------------|-----------------|----------------|
| Normal - Difficult | 25.89           | < .001         |
| Normal - Fast      | 17.33           | < .001         |
| Difficult - Fast   | -8.57           | 0.005          |

**Table 6 Condition Pair Comparison in SI***Condition Pair Comparison in SI*

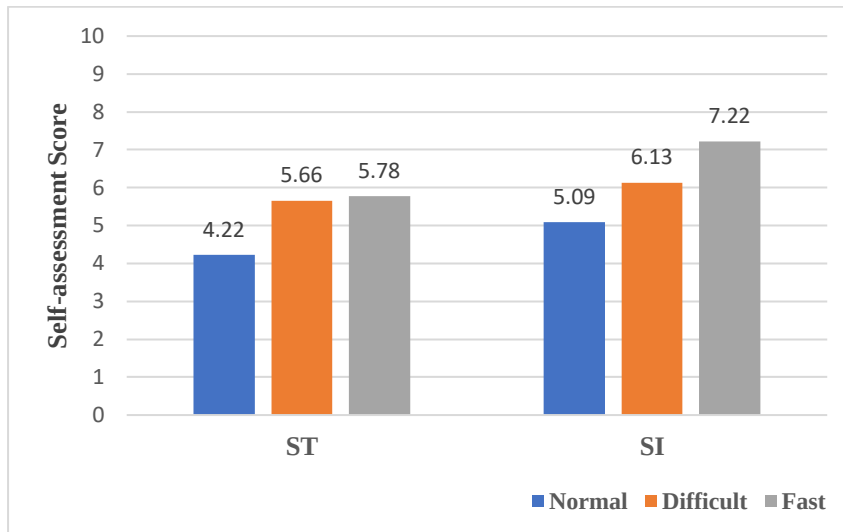
| Condition Pair     | Mean Difference | <i>P</i> value |
|--------------------|-----------------|----------------|
| Normal - Difficult | 16.66           | < .001         |
| Normal - Fast      | 26.80           | < .001         |
| Difficult - Fast   | 10.14           | < .001         |

#### 4.5.2 Subjective Analysis: Perceived Cognitive Load Results

This section presents the results of the self-assessment. Participants perceived cognitive load was measured on a 10-point scale, ranging from 1 to 10, based on the NASA Task Load Index. An overview of the scores for both SI and ST tasks under the three conditions can be found in Figure 6. Specifically, for ST, the mean perceived cognitive load self-assessment rates in the three different conditions were as follows: Normal (Mean = 4.22; SD = 2.2), Difficult (Mean = 5.66; SD = 2.13), Fast (Mean = 5.78, SD = 2.27). For SI, the outcomes were: Normal (Mean = 5.09, SD = 2.08), Difficult (Mean = 6.13; SD = 2.15), Fast (Mean = 7.22; SD = 2.56).

**Figure 6 Mean Perceived Cognitive Load Scores**

*Mean Perceived Cognitive Load Self-assessment Scores in ST and SI*



The repeated measures ANOVA on perceived cognitive load self-assessment scores showed that the effect of task was significant ( $F(1, 31) = 11.29, p = .002$ ), showing that participants' perceived cognitive load level in the ST task (Mean = 5.22, SD = 2.29) was lower than in the SI task (Mean = 6.15, SD = 2.41), the mean difference is -0.93. The interaction between each condition was also significant ( $F(2, 31) = 33.76, p < .001$ ), this result tells us that the differences between conditions (Normal, Difficult and Fast) are not the same for each task. The interaction between Tasks and Conditions was significant ( $F(2, 31) = 3.76, p = 0.03$ ), this result indicates that the three conditions in SI and ST task was different. The further Post-hoc comparisons of the three conditions showed that in ST, trainees' perceived cognitive load in Normal condition was lower than in both Difficult and Fast condition, and the mean difference were -1.44 and -1.56. There was no significant difference between Difficult condition and Fast condition. While in SI, participants perceived cognitive load in Normal condition was lower than that in both Difficult condition (Mean Difference = -1.04) and Fast condition (Mean Difference = -2.13). The perceived cognitive load in Difficult condition was lower than that in Fast condition (Mean Difference = -1.09). see Table 7 and Table 8.

**Table 7 Condition Pair Comparison in ST***Condition Pair Comparison in ST*

| Condition Pair     | Mean Difference | <i>P</i> value |
|--------------------|-----------------|----------------|
| Normal - Difficult | -1.44           | 0.002          |
| Normal - Fast      | -1.56           | 0.003          |
| Difficult - Fast   | -0.12           | 0.987          |

**Table 8 Condition Pair Comparison in SI***Condition Pair Comparison in SI*

| Condition Pair     | Mean Difference | <i>P</i> value |
|--------------------|-----------------|----------------|
| Normal - Difficult | -1.04           | < .001         |
| Normal - Fast      | -2.13           | < .001         |
| Difficult - Fast   | -1.09           | 0.001          |

#### 4.5.3 The Relationship between Accuracy and Perceived Cognitive Load

In the online study, we were looking to see if there is a relationship between participant's perceived cognitive load level and their accuracy scores. The correlation results showed that participant's accuracy scores were negative correlated with their perceived stress level ( $r(30) = -0.284$ ,  $p < .001$ ), showing that the higher the perceived cognitive load score, the lower the accuracy score, and vice versa.

**Table 9 Correlation Result between Accuracy Score and Self-assessment Score***Correlation Result between Accuracy Score and Self-assessment Score*

|                               | 1         | 2 |
|-------------------------------|-----------|---|
| 1. Mean Accuracy Score        | —         |   |
| 2. Mean Self-assessment Score | -0.284*** | — |

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

#### 4.5.4 Working Memory Capacity Results

A correlation test was conducted between working memory capacity and accuracy score, and the results showed that there was no significant correlation between them ( $r(30) = -0.06$ ,  $p = 0.42$ ). The table below shows the correlation results.

**Table 10 Correlation Result of Working Memory Capacity**

*Correlation Result of Working Memory Capacity*

|                        | 1    | 2 |
|------------------------|------|---|
| 1. Mean Accuracy Score | —    |   |
| 2. WMC Score           | 0.42 |   |

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

#### 4.6 Discussion

I conducted a correlation test between the results of self-assessment and accuracy. The primary motivation behind conducting a correlation test was to validate the consistency between these two evaluation methods. A negative correlation between the two would imply that as trainees report a higher perceived cognitive load, their task performance (in terms of accuracy) tends to diminish. This would be in line with theoretical expectations. Confirming this hypothesized relationship not only strengthens the validity of each measure individually but also provides a comprehensive perspective on the cognitive load experienced by the participants.

In answering the research **Question 1 – ‘Do text difficulty and delivery rate influence trainees’ cognitive load in SI tasks?’** Research using performance method indicates that the results for SI accuracy revealed that the highest accuracy score was observed under the Normal conditions. Conversely, the accuracy score under the Fast condition was lower than the score under the Difficult condition. Research using subjective method show that in the SI task, the highest perceived cognitive load was noted under the Fast condition. The perceived cognitive load under the Difficult and Normal conditions followed, ranking second and third respectively. These findings indicate that both text difficulty and delivery rate exert an influence on trainees’ cognitive load during SI, and the impact of delivery rate appears to be more profound. Indeed, simultaneous interpreting between Chinese and English already constitutes a significant

cognitive challenge for trainee interpreters. When there is a substantial grammatical disparity between the source and target languages, it amplifies the demand on processing capacity within short-term memory. This heightened demand can precipitate cognitive saturation, forcing interpreters to store large amounts of information temporarily before they can articulate it effectively in the target language (Li 2001; Gile 2009; Ye and Shi 2009). In this context, under the Fast condition, the additional cognitive load imposed by the high speech rate compounds the cognitive burden induced by the disparities in language systems, thereby impinging on output performance. Moreover, Gile (2009) proposed a link between high speech density and speech rate, arguing that an increased density of the source speech intensifies processing capacity requirements. Thus, he classifies this phenomenon under “cognitive problem triggers” (Gile, 2009, p. 192), which can cause challenges in SI task. Contrastingly, issues prompted by text difficulty seem to align more with “language-specificity related problems”, as outlined by Gile (2009, p. 194). These problems could encompass factors such as differences in the perception of words, grammatical redundancies, syntactic structures, among others. According to the findings in Experiment One, it appears that cognitive problem triggers have a more pronounced effect on the interpreter’s performance than the issues provoked by language-specificity related problems, thereby diminishing their interpreting accuracy. Continuing this line of thought, it becomes evident that trainees need to focus on enhancing their cognitive coordination abilities in order to process more information under the Fast condition. Mismanagement of processing capacity could potentially lead to an increase in errors. Therefore, effective strategies for managing cognitive resources are crucial for reducing the cognitive load in the Fast condition. Regarding the elaboration of these results, I shall resume the discussion upon integrating the findings from the lab study with insights derived from participants’ interviews.

Building upon the exploration of my research, the following discussion will address the **Question 2 – ‘Do text difficulty and delivery rate influence trainees’ cognitive load in ST tasks?’** The results show that trainees also achieved highest accuracy score in the Normal condition during ST. Intriguingly, an opposite trend emerged compared to SI - the accuracy score for the Difficult condition was lower than that for the Fast condition in ST. In addition, trainees perceive a higher cognitive load in both the Difficult condition and Fast condition compared to their perceived cognitive load in the Normal condition. However, no significant difference was discerned in the perceived cognitive load scores between the Difficult and Fast conditions. On one hand, the significant impact of text difficulty could be attributed to the fact

that when sentences are not lengthy or are composed of independent clauses, trainee interpreters can employ a strategy of sequential driving to divide the text into translatable units. However, as sentences become lengthier (sentence length was one of the indicators used to measure text difficulty in this study), trainee interpreters need to read more than one unit to rephrase the sentence. This not only increases the burden on their memory but also causes deviation in their coordination of cognitive management. On the other hand, these results appear to affirm Gile's (1995) assertion that during the Reading Effort in sight translation, translators lack the vocal cues such as intonation, hesitations, or other pauses. These cues, available in simultaneous and consecutive interpreting, facilitate the segmentation of the text into translation units, thereby heightening the complexity of the ST task. Meanwhile, even though trainees can self-regulate their pace under the Fast condition, the time pressure still elevates their cognitive load. Some studies (Jensen and Jakobsen 2000; Hansen 2006; Sharmin et al. 2008; Ghobadi et al. 2017) have explored the relationship between time pressure and written translation. However, there has always been a lack of empirical evidence regarding whether time pressure would impact sight translation task. The increased delivery rate seems to intensify trainees Memory Effort, particularly if they are trying to keep pace with quickly moving source text. Time pressure may also limit the use or effectiveness of certain cognitive or translation strategies that the trainees might otherwise employ.

Proceeding with a detailed discussion of the difference between the two tasks - **Question 3 – ‘Do trainees’ cognitive load differ between the ST and SI tasks?’** Research using performance method indicates that students achieved higher scores in terms of accuracy in the ST compared to their accuracy in the SI under all three conditions of Normal, Difficult, and Fast tasks. Research using subjective method show that the average perceived cognitive load of students under all three conditions in the ST task was lower than that of students under all three conditions in the SI task. Combining these two results, it can be inferred that the cognitive load experienced by trainees during simultaneous interpreting is greater than that experienced during sight translation. The performance result came as no surprise and could be found in earlier studies (Viezzi 1989; Lambert 2004). The overall cognitive load experienced by trainees during SI task is generally higher than during ST task, a difference that can likely be attributed to distinct input modes and information processing strategies inherent to each task type. As proposed by Chernov (2004), the cognitive load imposed on a simultaneous interpreter is substantial, necessitating real-time anticipation of the speaker's statements for successful interpretation. This constant predictive cognitive effort is paramount for SI task, significantly

contributing to the high cognitive load. Contrastingly, in ST tasks, interpreters have the advantage of visually accessing the entire sentence in advance, as opposed to SI tasks where they must process auditory information in real-time. This differing approach reduces the cognitive effort dedicated to prediction, potentially mitigating the overall cognitive load. Moreover, the burden on short-term memory appears to be less in sight translation compared to simultaneous or consecutive interpreting. It seems that the inherent characteristics of these tasks seems to influence the magnitude of cognitive load experienced by the interpreters. Another reason contributing to the lower cognitive load experienced by trainee interpreters during ST tasks is that ST is self-paced, in contrast to SI, which is paced by the speaker of the source language. Even though the current study imposes time constraints on ST tasks, with each sentence displayed on the screen for the same duration as in SI tasks, subjects can still regulate their pacing. This self-regulation of pace offers a significant advantage, allowing interpreters to control their rhythm during the process of the tasks.

Another interesting finding is that under the Difficult condition, trainees' accuracy scores in sight translation were a mere 3.81 points higher than those in simultaneous interpreting, and the difference was not statistically significant. This may imply that when dealing with difficult texts, trainee interpreters/translators may not necessarily confront more challenges in the 'auditory' mode than they do in the 'visual' one. Although this outcome does not entirely align with the findings reported by Jereščenková (2014) and Chmiel et al. (2020), which suggests that ST has more lexical interference than SI, it does imply that trainees might need to manage more source text interference during the ST task under the Difficult text condition. Furthermore, during the ST task, the source language remains continuously visible to the trainees, despite their gaze focusing only on a small text segment at a time. This presence persists even when we attempted to reduce the visual interference of the text in the experiment by presenting it sentence by sentence. The enduring visibility of the source language heightens the possibility of interference between the two languages, which may affect ST performance (Agrifoglio 2004; Gile 2009; Shreve 2010). This situation poses a substantial challenge for trainee interpreters and trainee translators, and they may need more training or learn more anti-interference strategies to overcome the visual interference.

**Question 4 – ‘Does working memory capacity, when considered as a covariate variable, correlate with the accuracy of interpreting trainees in SI and ST tasks?’** We included working memory capacity (WMC) as a covariate in our study. A covariate, by its nature, is a

type of control variable that could potentially influence the outcome of an experiment but is not the primary variable of interest. Given the significant role WMC plays within cognitive load study and considering the evidence of its impact on interpreting performance in existing literature, we felt it was important to account for WMC in our investigation. However, our findings from the online study did not reveal a significant correlation between WMC and the accuracy scores of our participants. This may suggest that WMC, despite its general importance to cognitive tasks, may not be a decisive factor influencing interpreting accuracy in this specific context. This result does not diminish the potential role of working memory in interpreting tasks in general. Instead, it suggests that when considering individual differences in SI and ST accuracy observed in Experiment One, the potential influence of variations in individual WMC might not necessarily need to be taken into account. Nevertheless, it is also worth considering the testing method used for WMC. The Corsi Span Test, a popular measure of measuring WMC, was employed in Experiment One. It might not be the most suitable method for SI and ST task, as interpreters' or translators' memory spans may be more reliant on their word span. Word-list stimuli could potentially offer a more appropriate measure, as supported by studies examining the relationship between interpreters' WMC and their performance (Darò and Fabbro 1994; Christoffels et al. 2006; Köpke and Nespoulous 2006; Chmiel 2018). Therefore, in the subsequent lab experiment (Experiment Two), I utilized two versions of a wordlist/word span test instead of the Corsi test, in the hope of generating more accurate insights into the correlation between WMC and interpreting performance.

#### **4.7 Limitations and Implications**

While this online study provides valuable insights, there are notable limitations that should be acknowledged. These limitations, however, also present opportunities for refinement in the design of Experiment Two. First, the uncontrolled nature of the online experimental environment presents a significant challenge. While online experiments offer notable benefits such as enabling widespread access to participants worldwide, they are not without drawbacks. A major limitation is the lack of a uniform experimental environment for all participants, as each participant takes part in the experiment in various settings. In contrast, lab experiment is conducted under standardized conditions, ensuring consistent environments for all participants. Recognising this limitation underscores the importance of a controlled environment in obtaining reliable results.

Second, the source text selection process in the online study necessitates refinement. The selection process for participants was also more stringent. In order to enhance the participants' familiarity with the subject matter, the subsequent lab experiment used texts related to the Covid-19 pandemic, a topic of significant global relevance at the time of the study. Moreover, to align the experiment closer to their regular training, the selected texts focused on the discussion of the economic and political impacts of the pandemic. This strategic choice of topic and text style aims to provide a more familiar context for the participants, thereby making the experiment closer to their usual study material.

Third, the counterbalancing of the experimental design requires improvement. Due to time constraints, the online study was kept brief, thus excluding a 'Difficult Fast' condition. However, to ensure a more balanced design, this condition will be integrated into the lab experiment, increasing the breadth of the data collected.

Lastly, although perceived cognitive load self-assessment method is commonly used in some cognitive load related studies, their subjective nature can potentially undermine the reliability of results. Consequently, the lab experiment will endeavour to measure cognitive load more objectively by utilising physiological markers such as pupil dilation and galvanic skin response (GSR). These measures offer a more reliable approximation of cognitive load, thus significantly enhancing the rigor of the study. Through acknowledging and addressing these limitations, the design of the forthcoming laboratory experiment will be significantly improved, increasing the potential for more reliable findings.

## Chapter 5 Experiment Two

The chapter begins with an overview of the research design, segueing into details about participant backgrounds and the nature of the experimental stimuli. Further sections elaborate on the research instruments employed and the procedural steps of the experiment. The processes of data collection and its subsequent treatment are addressed thereafter. Later sections of the chapter spotlight the results, focusing on elements like accuracy, cognitive load, among other crucial variables. The discussion portion offers a deeper analysis of the findings, emphasizing aspects such as delivery rate and text complexity, and aligns them with existing literature for a comprehensive understanding.

### 5.1 Research Questions, Research Variables and Hypothesis

#### 5.1.1 Research Questions

**Question 1:** How does text difficulty and delivery rate impact on trainees' cognitive load during an SI task?

**Question 2:** How does text difficulty and delivery rate impact on trainees' cognitive load during a ST task?

**Question 3:** How does cognitive load differ between ST and SI tasks among interpreting trainees?

**Question 4:** Does the impact of text difficulty and delivery rate on cognitive load differ between ST and SI tasks performed by interpreting trainees?

#### 5.1.2 Research Variables

- Independent Variables:
  - a) Task with the levels: sight translation (ST) and simultaneous interpreting (SI)
  - b) Delivery rate with the levels: normal vs. fast
  - c) Text difficulty with the levels: normal vs. difficult
- Dependent Variables:
  - a) ST and SI performance accuracy score
  - b) Pupil size
  - c) GSR
- Covariate Variables: Working Memory Capacity (WMC)

### **5.1.3 Hypothesis in Experiment Two**

H1: Both text difficulty and delivery rate impact on trainees' cognitive load during an SI task.

H2: Both text difficulty and delivery rate impact on trainees' cognitive load during an ST task.

H3: Trainees experienced higher cognitive load in SI than in ST task.

## **5.2 Research Design**

The Experiment Two has been carried out in a laboratory of the Psychology Department at the University of Stirling. The Psychology Department provided the eye tracker and skin response transducer. In order to make the experimental environment as close as possible to the real interpreting working environment, we used baffles to build a simple interpreting booth and equipped with headphones and microphones. The within-subject design is also applied to this lab experiment. The following part is about the experimental design.

### **5.2.1 Experiment Instruments**

#### **5.2.1.1 Eye tracker**

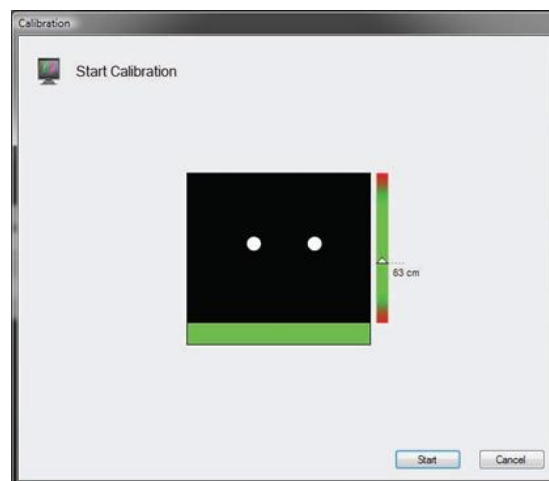
In the current experiment, stimuli were presented using a Tobii 1750 eye tracker (see figure 7), integrated into a 17-inch monitor with a 60Hz refresh rate. The presentation was facilitated using E-prime software. This Tobii system, positioned approximately 64 cm from the participant, not only offers a comprehensive calibration sequence with different point options but also boasts an accuracy of approximately 0.5 degrees of visual angle. Its unique design ensures consistent camera positioning relative to the display and offers a wide viewing angle, making additional tracking adjustments unnecessary. Furthermore, the Tobii 1750 is equipped with a robust array of infrared (IR) LEDs for scene illumination and employs a dual-camera system to simultaneously track both eyes, capturing a broad range of gaze angles and accommodating minor head movements.



**Figure 7 The Tobii 1750 Eye-tracker**

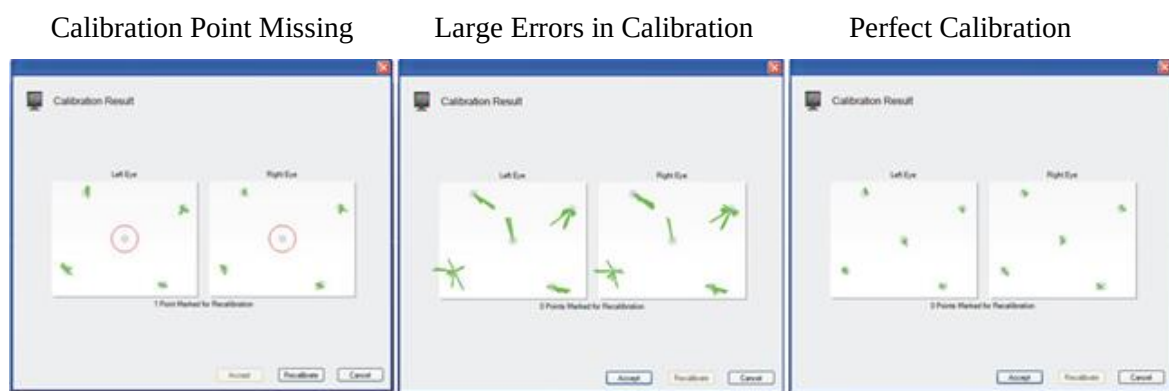
### **5.2.1.2 Eye-tracker Calibration Settings**

In preparation for the session of the pupil size data collection, participants undergo a calibration process. This essential step enables the eye tracker to capture specific characteristics of the participant's eyes. The calibration process utilizes both the recorded eye attributes and an internal, physiological 3D model of the eye. During the first calibration phase, we use track status box to help adjust the participants' sitting position and the eye tracker position. In the tracking status box, the participants' eyes are represented as two white dots and should be centrally located. The distance indicator must register a value ranging from 50 to 80 cm, and the bar at the track status box's base should exhibit a green colour (see Figure 8).



**Figure 8 Correct Eye Positioning in the Track Status Box**

During the next calibration phase, participants were asked to look at the designated points as they move over the screen. As participants engage with these points, multiple eye images are captured and analysed. The calibration procedure in Tobii Studio in the Experiment Two presents five calibration dots against a uniform background colour. Once all calibration points have been displayed, the calibration process will conclude on its own. After completion, the results can be viewed in two vector plots within the Calibration Result dialog box. The calibration plot displays error vectors using green lines. The length of each green line represents the discrepancy between the gaze point determined by the eye tracker and the actual position of the dot.



**Figure 9 Calibration Results**

### 5.2.1.3 E-Prime for Tobii

In Experiment Two, the E-Prime Extensions software was utilized to design the experiment, specifically for presenting the stimuli (e.g., present the source text/speech). E-Prime Extensions for Tobii is a set of software extensions to E-Prime, enabling it to interact with the Tobii eye tracker Server during an experiment's execution. It allows the design of new E-Prime experiments or the modification of existing ones to work with the Tobii Eye Tracker. Additionally, the extensions also come with a collection of sample experiments that can either be directly executed or used as a foundation for creating new experiments.

### 5.2.1.4 GSR Device

GSR data was recorded by using the Biopac MP36. The Biopac system offers both wired and wearable wireless tools for physiology measurement, facilitating in-depth in-lab exploration of cognitive functions. To record the skin response, we used the SS3L GSR transducer. The transducer was placed on the index and middle finger of the participants' hand (see figure 10).

The sensor is placed on the bottom of the participant's fingertip and wrapped around the finger with the Velcro tape. It is vital that the transducer fits snugly to the finger without being so tight that it cuts off blood circulation (Biopac Student Lab 2014). Due to its importance, I inquired about participants' comfort each time after they wore the transducer.



**Figure 10 “Correct setting of GSR” (Biopac Student Lab, 2014, p. 6)**

The MP35 must be connected to a laptop for operation. The experiments were executed using the Biopac Student Lab software installed on this laptop. The system's built-in microprocessor handles data acquisition and interaction with the laptop. The MP35 unit converts incoming signals into digital format, making them suitable for processing on the computer. The MP35 has four analog input channels, with one serving as a trigger input. The transducer also needs to connect to the MP35 device. It is crucial to ensure the input channels are correctly selected before starting an experiment. For this particular experiment, the chosen input was 'Channel 3 (CH3)'. The subsequent figure illustrates the interface presented on the laptop via the Biopac Student Lab once the MP35 device was connected.

### **5.2.2 Participants**

Before the lab experiment, information about participant recruitment was also posted on my personal social networking sites (WeChat Moments and Instagram). I also forwarded the recruitment notice to students who graduated from the University of Stirling and then went on to do master's programmes at different UK universities and asked them to share it on their social networking platforms. After registering for the experiment, participants will receive a brief pre-participation questionnaire to assess their background and determine their suitability for the study. As a supplement to the recruitment information, the pre-participation questionnaire aimed to gather basic information on participants' academic backgrounds and training experiences. This included details about the university they were currently enrolled in, as well as the duration they had studied English before starting their postgraduate translation

programme. Furthermore, the questionnaire inquired about any formal training they might have received in sight translation or simultaneous interpreting theories and practices. This information was crucial to ensure that all participants shared similar learning experiences and educational standards.

A total of twenty-two English-speaking Chinese students participated in the Experiment Two, and all of them were paid. Their ages also ranged from twenty-two to twenty-five years old, and three were male and nineteen were female. As reiterated from Experiment One, even though there were more female participants than male ones, the similarities in their personal backgrounds and training experiences likely offset the potential effects of the gender imbalance. A profile of the participants is provided below.

**Table 11 Profile of Participants**

|                                  |                                                                |
|----------------------------------|----------------------------------------------------------------|
| Total Participants               | 22 trainees                                                    |
| Age Range                        | 22—25                                                          |
| Current Degree Program           | MA/MSc Translation and Interpreting (UK University)            |
| Undergraduate Degree             | Bachelors in Translation and Interpreting Studies              |
| Native Language                  | Mandarin Chinese                                               |
| Secondary Language               | English                                                        |
| Age of Starting English Learning | 10-12 years                                                    |
| Training in SI & ST              | Received at least 10 weeks of relevant master course training. |
| Professional Experience          | None in interpreting or translation at the time of experiment  |

### 5.2.3 Experimental Stimuli

#### 5.2.3.1 Brief on Source Texts

Both ST and SI tasks were done from English into Chinese. All of the texts used in Experiment Two were excerpts from video transcripts, articles and websites. Due to the addition of a variable condition, a total of eight texts used in Experiment Two. Each text contains either 127

or 128 words and consists of five sentences. The order of the source texts was counterbalanced across participants. The eight source texts were used in both the SI and ST tasks, but participants received different texts for each task. This means the ST texts done by some participants also served as the SI (video) done by other participants, while the SI texts (video) done by some participants also served as the ST texts done by other participants. The eight texts were set as follows:

- Two texts of **Normal Difficult Text with Normal Delivery Rate**
- Two texts of **Difficult Text with Normal Delivery Rate**
- Two texts of **Normal Difficulty Text with Fast Delivery Rate**
- Two Texts **Of Difficult Text with Fast Delivery Rate**

#### **5.2.3.2 Evaluation of Source Text Difficulty**

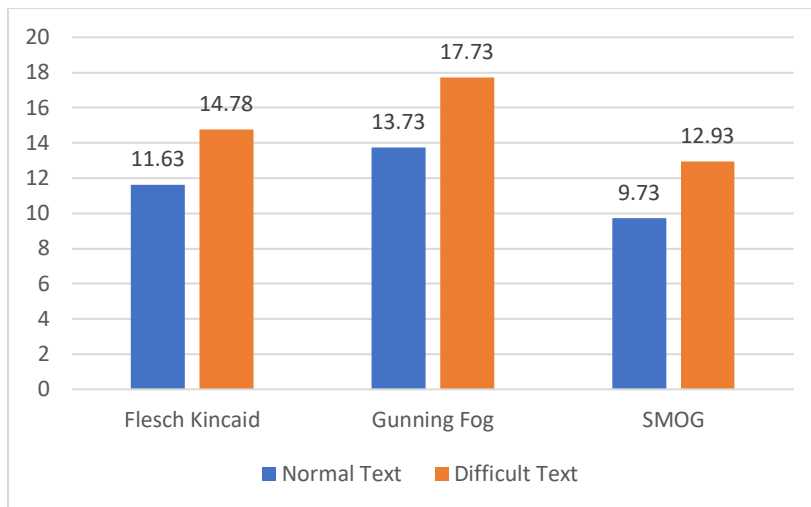
The Experiment Two employed both objective/quantitative and subjective/qualitative measures to evaluate the difficulty level of the source texts. From a readability standpoint as one of the quantitative measures, the study incorporated four mainstream metrics, namely, Flesch Reading Ease, Flesch Kincaid Grade Level, SMOG Index, and Gunning Fog Score. These metrics are frequently utilized for gauging the readability of texts. Moreover, the Word Frequency Index was also adopted as a standard in the quantitative evaluation of text difficulties. On the other hand, both readability formulas and word frequency, as previously noted, have their inherent limitations. In Experiment Two, I made improvements to the text difficulty assessment approach by incorporating qualitative methods. Specifically, bilingual individuals were engaged to evaluate the difficulty of the texts. This additional step ensured a more comprehensive assessment and comparability between the Normal and Difficult texts.

##### **5.2.3.2.1 Quantitative Measure: Readability Measurement**

As mentioned before, the Flesch Reading Ease, Flesch Kincaid Grade Level, and SMOG Index are frequently utilized metrics in the field of translation and interpreting studies, especially when managing text complexity. Conversely, the Gunning Fog Score aims to gauge the years of formal education needed for someone to understand a text at first glance. Average scores for both Normal and Difficult texts have been provided below:

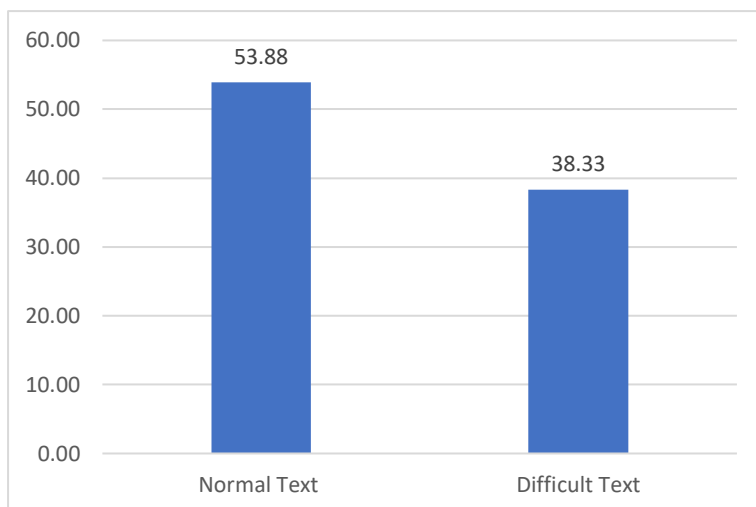
**Figure 11 Readability Test Accuracy Scores**

*Average Scores of Flesch Kincaid Grade Level, Gunning Fog and SMOG Index of Four Normal Texts and Four Difficult Texts*



**Figure 12 Readability Test Accuracy Scores**

*Average Score of Flesch Reading Ease of Four Normal Texts and Four Difficult Texts*



In this study, the application of the Flesch Kincaid, Gunning Fog, and SMOG indexes for readability assessment revealed a distinct escalation in complexity levels from Normal to Difficult text, as illustrated by the figures. The Flesch Reading Ease index assigns numerical scores ranging from 0 to 100. Higher scores are indicative of a text's ease of understanding, while lower scores suggest that a text is harder to comprehend. From the data presented, it can be observed that the average score for the Normal texts was 53.88, and for the Difficult texts,

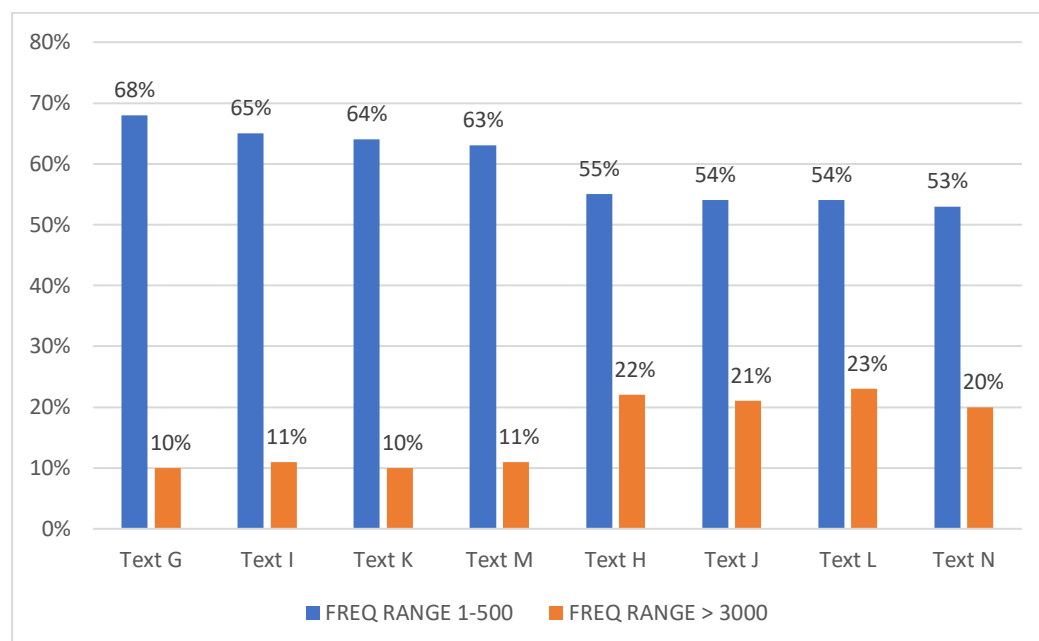
it was 38.33. According to Flesch (1948), scores ranging from 30 to 50 are categorised as ‘difficult’.

### 5.2.3.2.2 Quantitative Measure: Word Frequency Measurement

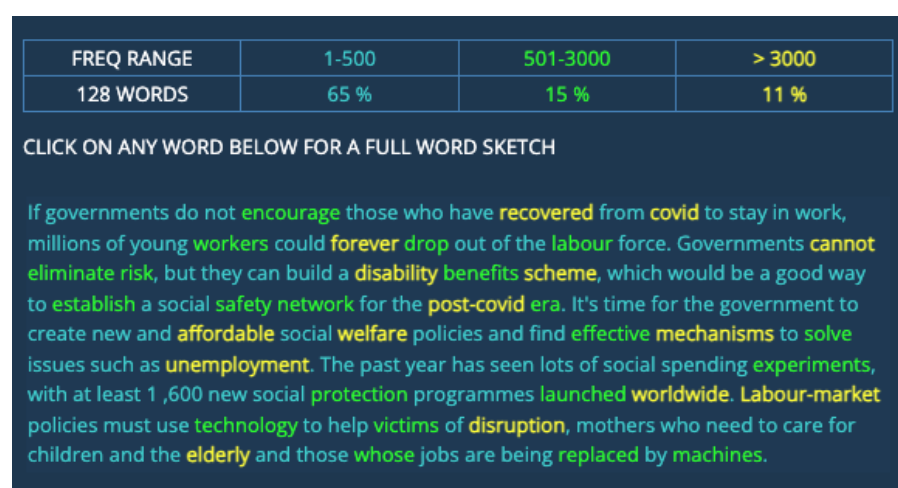
The Word Frequency measurement was also incorporated into the present study. Drawing on the approach of Wang (2017), the tool (<http://www.wordandphrase.info/frequencyList.asp>) used to calculate the word frequency of each text. Here, words from the source texts were categorized into high-frequency and low-frequency groups. The criteria for these classifications are as follows: High-frequency words rank within the 1-500 range, whereas low-frequency words rank above 3000. In the subsequent figure, a comparison is made between the count of low-frequency words and that of high-frequency words within both the Normal texts (Text G, Text I, Text K, Text M) and Difficult texts (Text H, Text J, Text L, Text N):

**Figure 13 Word frequency scores**

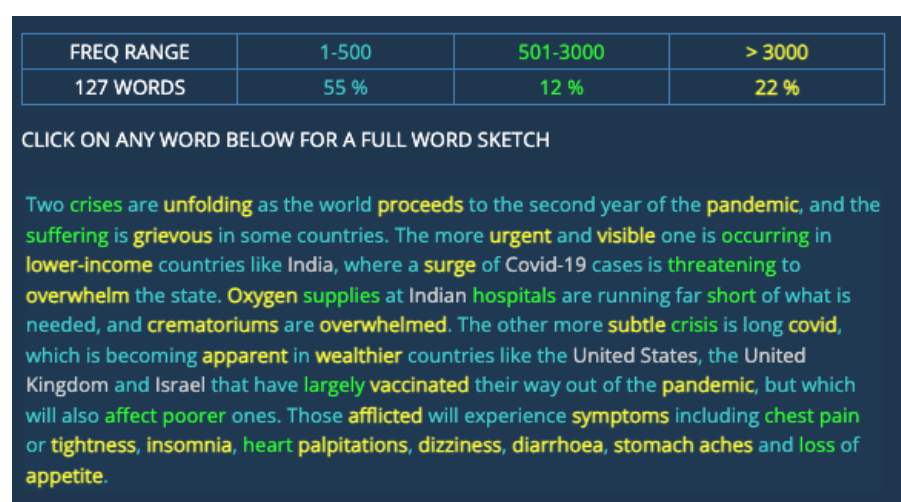
*Word frequency scores of Each Text*



**Figure 14 Word Frequency Evaluation of Normal Text I**



**Figure 15 Word Frequency Evaluation of Difficult Text H**



### 5.2.3.2.3 Qualitative Measure: Bilinguals' Rating

Five female bilingual evaluators were invited to evaluate the difficulty of source texts. Three of them were native Chinese speakers, while the other two were native English speakers. All five evaluators have educational backgrounds in Translation and Interpreting Studies, and their profiles are provided in Table. Importantly, none of these raters were experiment subjects in the lab study. The raters were tasked with evaluating the texts from three dimensions: word difficulty, syntactic difficulty, and reading difficulty. They were asked to rate on a 5-point scale, ranging from 1 (indicating 'very easy') to 5 (indicating 'highly difficult').

**Table 12 Five Bilinguals' Profile**

| Raters      | Age | Native Language | Second Language | Education Background | Notes                                                                                         |
|-------------|-----|-----------------|-----------------|----------------------|-----------------------------------------------------------------------------------------------|
| Bilingual 1 | 29  | English         | Chinese         | MSc                  | MSc in Chinese-English Translation Studies; 2 years of Chinese-English translation experience |
| Bilingual 2 | 27  | English         | Chinese         | MA                   | MA in Chinese-English Translation Studies; 3 years of Chinese-English translation experience  |
| Bilingual 3 | 24  | Chinese         | English         | MA                   | BA with a background in Translation and Interpreting                                          |
| Bilingual 4 | 29  | Chinese         | English         | PhD                  | MA in English Studies; PhD focusing on Translation Studies                                    |
| Bilingual 5 | 24  | Chinese         | English         | MA                   | BA with a background in Translation and Interpreting                                          |

As mentioned before, the bilingual raters evaluated the texts based on three criteria: word difficulty, syntactic difficulty, and reading difficulty. The average scores from these three aspects are displayed in Table 13. To examine the significance of differences in scores evaluated by five raters to each text, an ANOVA was employed (see Table 14). The statistical analysis revealed no significant difference between texts classified as Normal, which is the same as the findings for Difficult texts. However, a significant difference was observed when comparing each Normal text with each Difficult text. This outcome further contributes to our understanding and evaluation of source text difficulty levels in the present study.

**Table 13 Bilingual Rater's Evaluation of Difficulty of Source Texts**

| Rater | Text G | Text H | Text I | Text J | Text K | Text L | Text M | Text N |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1     | 1.00   | 3.00   | 2.00   | 2.67   | 2.00   | 3.00   | 2.00   | 3.33   |
| 2     | 2.00   | 3.67   | 2.00   | 3.67   | 2.33   | 3.33   | 2.33   | 3.67   |
| 3     | 1.33   | 3.00   | 1.67   | 3.67   | 1.33   | 3.33   | 1.33   | 3.33   |
| 4     | 2.00   | 4.00   | 2.00   | 3.33   | 2.00   | 4.00   | 2.00   | 4.00   |
| 5     | 1.00   | 3.00   | 1.00   | 3.33   | 1.67   | 3.00   | 2.00   | 3.67   |

A repeated measures ANOVA revealed a significant main effect of Difficulty on the scores,  $F(1, 4) = 155.19$ ,  $p < .001$ . This indicates that there was a significant difference in scores between the normal texts and the difficult texts. The following table provides a detailed comparison between the texts.

**Table 14 ANOVA: Post Hoc Comparisons – Text Type**

*ANOVA: Post Hoc Comparisons – Text Type*

| Text Type  | Text Type    | p-value | Significance | Text Type  | Text Type    | p-value | Significance |
|------------|--------------|---------|--------------|------------|--------------|---------|--------------|
| Text G (n) | - Text H (d) | < .001  | Yes          | Text I (n) | - Text K (n) | 0.986   | No           |
|            | - Text I (n) | 0.833   | No           |            | - Text L (d) | 0.01    | Yes          |
|            | - Text J (d) | 0.006   | Yes          |            | - Text M (n) | 0.974   | No           |
|            | - Text K (n) | 0.542   | No           |            | - Text N (d) | 0.01    | Yes          |
|            | - Text L (d) | 0.001   | Yes          | Text J (d) | - Text K (n) | 0.043   | Yes          |
|            | - Text M (n) | 0.54    | No           |            | - Text L (d) | 1       | No           |
|            | - Text N (d) | 0.002   | Yes          |            | - Text M (n) | 0.048   | Yes          |
| Text H (d) | - Text I (n) | 0.01    | Yes          |            | - Text N (d) | 0.839   | No           |
|            | - Text J (d) | 1       | No           | Text K (n) | - Text L (d) | 0.023   | Yes          |
|            | - Text K (n) | 0.008   | Yes          |            | - Text M (n) | 0.951   | No           |
|            | - Text L (d) | 1       | No           |            | - Text N (d) | 0.004   | Yes          |
|            | - Text M (n) | 0.016   | Yes          | Text L (d) | - Text M (n) | 0.036   | Yes          |
|            | - Text N (d) | 0.513   | No           |            | - Text N (d) | 0.509   | No           |
| Text I (n) | - Text J (d) | 0.039   | Yes          | Text M (n) | - Text N (d) | 0.003   | Yes          |

Concluding the section on text difficulty standards, it is worth acknowledging that the methodologies employed in this study were rather ambitious and may present certain limitations. However, the use of both quantitative and qualitative measures, including readability formulas and word frequency assessments, as well as bilingual evaluations, provided a multifaceted approach to text difficulty assessment. While the varying backgrounds of the evaluators may introduce some degree of variability, they also enriched the assessment by offering diverse perspectives. The indicators used in this study serve as evidence, highlighting the difficulty level of the source texts and shedding light on the features of text difficulty.

#### **5.2.4 Constraints on Delivery Rate and Presentation of Source Texts**

In Experiment Two, the delivery rate (including pause) settings were as follows:

- High delivery rate: 125 words per minute
- Normal delivery rate: 105 words per minute

Similar to the settings in Experiment One, in sight translation task, each sentence remained on the screen for a duration equivalent to the time taken by the speaker to speak that particular sentence. Each sentence in the sight translation task was individually presented in consecutive slides.

#### **5.2.5 Speaker in the SI Task**

The speaker for the SI tasks in Experiment Two is the same individual from Experiment One, ensuring consistency across both phases of the study. This person is a native English speaker who has always resided in English-speaking countries. Additionally, the speaker has a basic proficiency in oral Chinese communication, having lived in China between 2011 and 2012. To ensure the most appropriate recordings were used for the study, the speaker filmed the speeches in English, creating three different video versions for each source text. These recordings were then sent to me for final selection.

#### **5.2.6 Stimuli for Word Span Test**

In Experiment Two, both visual and auditory formats of the word span test were utilized. A total of 208 lowercase nouns were sourced from the MRC Psycholinguistic Database (Version

2.00) for the study. These were equally split between the visual and auditory WMC tests. The words were diverse in syllable count, spanning from one to four syllables. Additionally, each selected word consisted of four to seven letters. Within this Word Span Test, the word frequency was set between 100 to 400. The words for the auditory test were read by the same individual who served as the speaker in the SI task.

### **5.2.7 Fixed Questions in Semi-structured Interviews**

A ten-minute semi-structured interview was arranged for each participant. The interview aimed to gather qualitative data on participants' views of delivery rate and text difficulty—identified as potential triggers during the SI and ST processes. It also explored the reasons they might be influenced by these factors, their comparative views on the difficulty of SI versus ST, and the strategies they employed to handle difficult text and high delivery rates. The researcher translated the fixed interview questions for analysis, as the interviews with participants were conducted in Chinese.

## **5.3 Experiment Procedures**

Similar to online study, the order of the ST and SI tasks was counterbalanced across participants. Each participant was tested individually. The whole experiment lasts approximately thirty minutes and includes the following phrases: introduction, SI and ST task, WMC task and interview. All their performance during the experiment were recorded in audio format. The detailed procedures were as follows:

### **Phase 1: Setup and Introduction**

Participants were introduced to the procedure and were then given time to read and fill out the consent form.

### **Phase 2: Equipment Setup and Eye Tracker Calibration**

Participants were asked to wear the GSR equipment on their fingers. Subsequently, each participant's eye tracker was calibrated and validated to ensure accuracy before the test began.

### **Phase 3: Sight Translation and Simultaneous Interpreting Task**

Participants were prompted to sight translate and simultaneously interpret eight materials from English to Chinese. The sequence of ST and SI tasks was adjusted to ensure balance: some participants began with ST and then transitioned to SI, whereas others initiated with SI before

switching to ST. Moreover, to negate possible sequence biases, the arrangement of the four texts within each ST and SI task was shuffled randomly.

#### **Phase 4: Working Memory Capacity Task**

Upon completing the ST and SI tasks, participants transitioned to the word span test, which consisted of both visual and auditory word list recall sections. Visual Recall: Participants viewed eight groups of words on the screen, with each group consisting of two trials. The number of words gradually increased, starting with three words in the initial group and culminating with ten words in the final group. Auditory Recall: Participants listened to eight groups of words through headphones, with each group also consisting of two trials. Similarly, the groups of words increased in number, from three words in the beginning to ten words in the final group. The subjects were required to recall exactly all of each trial's words in the order they heard and saw them. They were told that the number of words in each group would increase over the test.

#### **Phase 5: Semi-structured Interview**

Concluding the tasks, participants entered the interview phase. Here, they answered questions regarding their thoughts on their performance during the tasks.

### **5.4 Data Treatment**

#### **5.4.1 Pupil Size Data and GSR Data**

Given the vast amount of mean pupil size data collected, we employed RStudio (Version 2021) for data processing. RStudio provides a flexible environment for data manipulation, helping to bridge any gaps and standardize the dataset for a more streamlined analysis. The Galvanic Skin Response (GSR) data was generated using the Biopac Student Lab software. After obtaining the initial data from Biopac, each participant's mean GSR readings were individually exported.

#### **5.4.2 SI and ST Accuracy Scoring**

Participants' SI and ST accuracy scores in the present experiment marked by the same markers who graded the accuracy scores in Experiment One. The assessment criteria and agreement are the same as in Experiment One. The marking was done by two experts who have been working as lecturers of Translation and Interpreting Studies for nearly 15 years. They both speak Mandarin Chinese as a first language and English as a second language.

In Experiment Two, the inter-rater reliability between two expert raters' assessment of participants' accuracy in SI and ST was also evaluated using pairwise correlation analysis. This analysis produced a correlation coefficient ( $r(30) = 0.994, p < .001$ ), demonstrating an high level of agreement between the raters' evaluations. Such a significant correlation coefficient underscores the considerable reliability of their markings for SI and ST accuracy.

### **5.4.3 Word Span Task Scoring**

Our WMC scoring method was similar to Pointe and Engle's (1990, p. 1120) "strict span score", which developed from Daneman and Carpenter's (1980) method. In the visual and auditory WMC test, the words were divided into eight groups and each group has two trials. One point was given if the subject perfectly recalled (according to the exact order) the words in two trials. If a subject perfectly recalled all the words in one trial but failed in the other, then half credit (0.5 points) was given. The level at which the subject was correct in eight groups was taken as a measure of their word span.

## **5.5 Results of Experiment Two**

This section reports the results of Experiment Two. All details of the whole results in Experiment Two can be found in Appendix. From now on, in order to make the description of the variables can be read clearly, I use:

- 'NN' to represent 'normal difficulty text, normal delivery';
- 'NF' to represent 'normal difficulty text, fast delivery rate';
- 'DN' to represent 'difficult text, normal delivery rate';
- 'DF' to represent 'difficult text, fast delivery rate'.

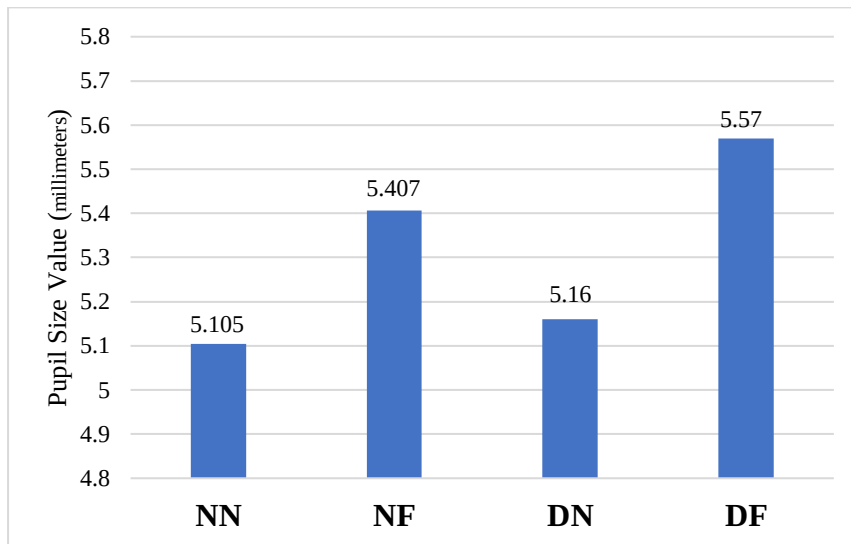
### **5.5.1 The Pupil Size Results**

The results of Mean pupil size data in SI and ST under four conditions were shown in Figure 16 and Figure 17. In SI task, participants' mean pupil sizes (in millimetres) for the four conditions were: NN (Mean = 5.105; SD = 0.78), NF (Mean = 5.407; SD = 0.74), DN (Mean = 5.160; SD = 0.76), and DF (Mean = 5.570; SD = 0.71). In ST task, the mean pupil sizes were: NN (Mean = 4.667; SD = 0.65), NF (Mean = 4.778; SD = 0.68), DN (Mean = 4.701; SD = 0.68), and DF (Mean = 4.750; SD = 0.66). Overall, the average pupil size for the SI task was

5.18 (SD = 0.78) and for the ST task was 4.72 (SD = 0.66), as illustrated in the accompanying Figure 18.

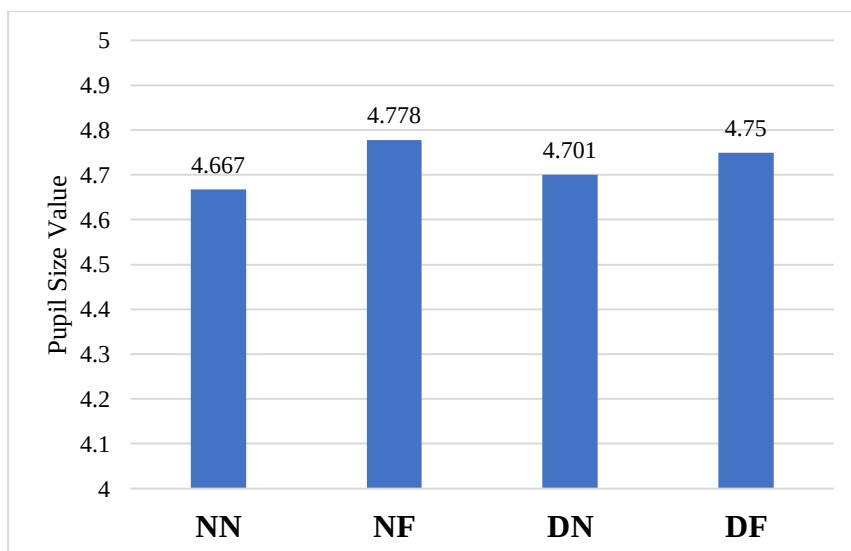
**Figure 16 Mean Pupil Size Value in SI**

*Mean Pupil Size Value under Four Conditions in SI*



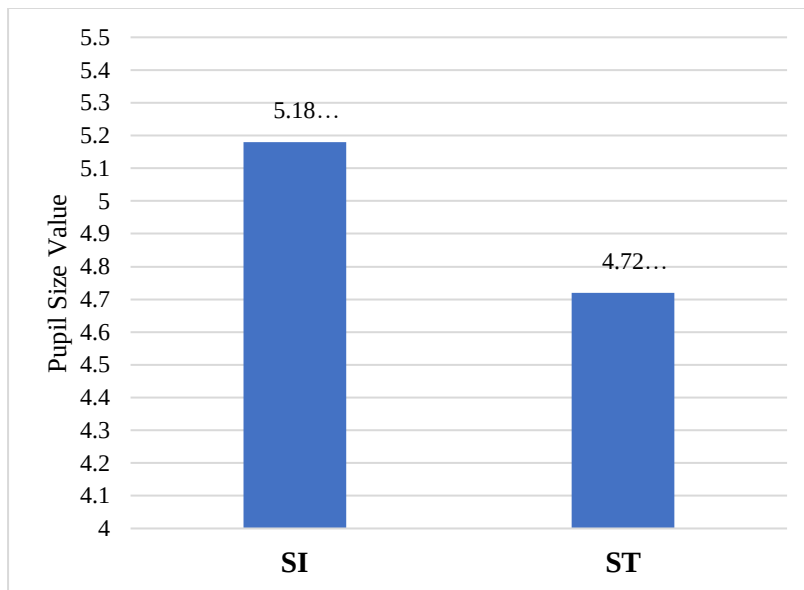
**Figure 17 Mean Pupil Size Value in ST**

*Mean Pupil Size Value under Four Conditions in ST*



**Figure 18 Mean Pupil Size Value in ST and SI**

*Mean pupil size value in SI and ST Task*



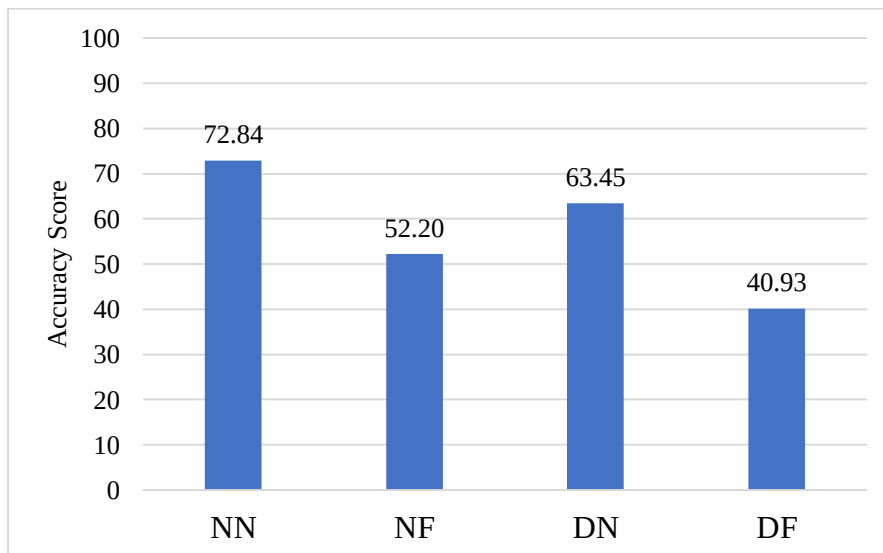
The further statistical analysis on pupil size data showed that the effect of the task was significant ( $F(1, 21) = 39.42, p < .001$ ), showing that participants' mean pupil size in SI task (Mean = 5.18; SD = 0.78) was higher than in ST task (Mean = 4.72; SD = 0.66). The repeated ANOVA showed that The interaction between Tasks and Difficulty was not significant ( $F(1, 21) = 0.65, p > 0.05$ ), and the effect of Difficulty was also not significant ( $F(1, 21) = 1.00, p = 0.33$ ), therefore, we can conclude that the text difficulty had no significant effect on participants' pupil size in either the ST or SI tasks. On the other hand, the interaction between Task and Delivery Rate was significant ( $F(1, 21) = 8.73, p < 0.05$ ), and the effect of Delivery Rate was also significant ( $F(1, 21) = 16.90, p < .001$ ). However, based on the further comparisons, in ST task, there was no significant difference between participants' pupil size in the 'normal' delivery rate condition (Mean = 4.68, SD = 0.65) and their pupil size in the 'fast' delivery rate condition (Mean = 4.76, SD = 0.68),  $t(21) = -0.86, p = 0.825$ . This suggests that the delivery rate variable does not significantly affect the trainee interpreters' pupil size in ST task. On the other hand, based on the further post hoc comparisons (Tasks \* Delivery Rate), in SI task, there was a significant difference between participants' pupil size in the 'normal' delivery rate condition and their pupil size in the 'fast' delivery rate condition,  $t(21) = -10.02, p < .001$ .

### 5.5.2 ST and SI Accuracy Results

Figure 19 and Figure 20 present the accuracy scores for ST and SI under four conditions. For SI, scores were: NN (Mean = 72.84; SD = 13.57), NF (Mean = 52.20; SD = 15.16), DN (Mean = 63.45; SD = 14.85), and DF (Mean = 40.93; SD = 9.58). For ST, they were: NN (Mean = 81.05; SD = 6.03), NF (Mean = 79.05; SD = 7.79), DN (Mean = 77.25; SD = 9.09), and DF (Mean = 73.68; SD = 9.69). Overall, the average accuracy score for the ST task was 77.76 (SD = 18.25) and for the SI task was 57.14 (SD = 8.40), as illustrated in the accompanying Figure 21.

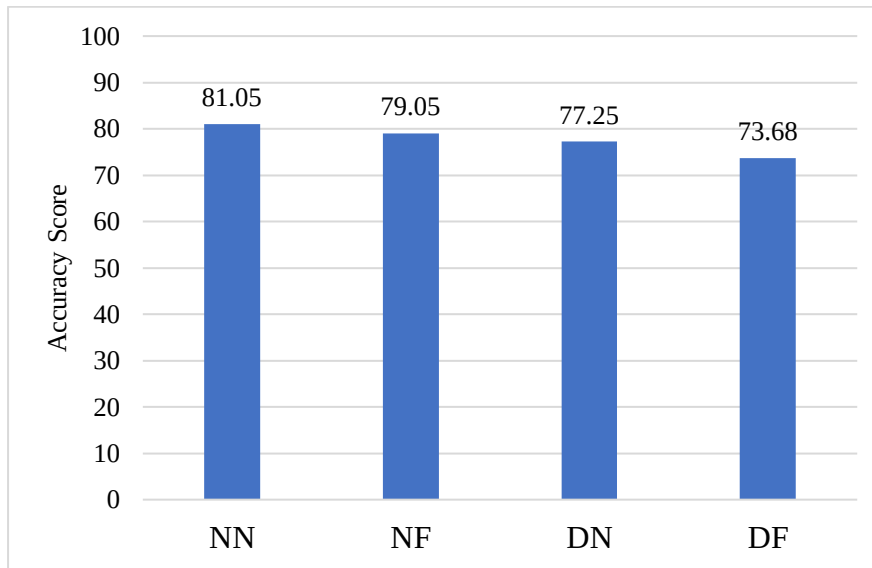
**Figure 19 Mean Accuracy Score in SI**

*Mean Accuracy Score under Four Conditions in SI*



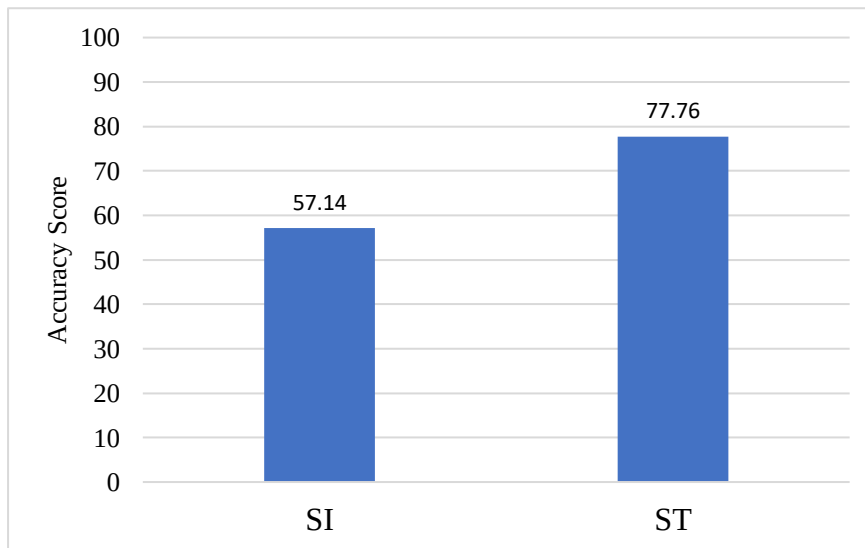
**Figure 20 Mean Accuracy Score in ST**

*Mean Accuracy Score under Four Conditions in ST*



**Figure 21 Mean Accuracy Score in SI and ST**

*Mean Accuracy Score of SI and ST*



The statistical analysis showed that the effect of the task was significant ( $F(1, 21) = 316.27, p < .001$ ), indicating that participants scored higher in the ST task (Mean = 77.76; SD = 18.25) than in the SI task (Mean = 57.14; SD = 8.40); Mean Difference = 20.62. In SI, the accuracy score under Normal condition was higher than under Difficult condition (Mean Difference = 10.33,  $t(21) = 13.63, p < .001$ ). The accuracy score under Normal condition was also higher than under Fast condition (Mean Difference = 21.58,  $t(21) = 22.58, p < .001$ ). In ST, the

accuracy score under Normal condition was higher than under Difficult condition (Mean Difference = 10.33,  $t(21) = 9.98$ ,  $p < .001$ ). The accuracy score under Normal condition was higher than under Fast condition (Mean Difference = 2.78,  $t(21) = 6.5$ ,  $p < .001$ ).

Considering the variations in text difficulty and delivery rate and their respective impacts on the accuracy of trainees in ST and SI tasks, it became imperative to delve deeper into which of these two variables exerts a more pronounced effect on accuracy across both tasks. Given the intricacies of drawing conclusions from raw scores between the variables text difficulty and delivery rate, a direct numerical comparison is not advisable. Instead of relying on raw values, I employed a Linear Mixed Model to ascertain the effect size. In this context, effect size refers to the average difference between a specific condition (e.g., NF, DN, or DF) and a baseline condition, which in this study is the NN condition. By comparing the values of the estimates, I can derive the following conclusions. In SI task (see Table 15), the estimate value of NF-NN being -20.66 implies that under the NF condition, students score on average 20.66 points lower than under the NN condition. This difference is mainly attributed to the increased speed of delivery. The estimate value of DN-NN being -9.39 suggests that, under the DN condition, students score on average 9.39 points lower than under the NN condition. This difference is primarily due to the increased text difficulty. Thus, the Linear Mixed Model result showed that in SI task, the variations in delivery rate have a more substantial impact on participants' output accuracy than text difficulty. On the other hand, in ST task (see Table 16), the estimate value of NF-NN being -2.00 implies that under the NF condition, students score on average 2.00 points lower than under the NN condition. The estimate value of DN-NN being -3.08 suggests that, under the DN condition, students score on average 3.08 points lower than under the NN condition. These results indicated that in ST task, the variations in text difficulty have a more substantial impact on participants' output accuracy than delivery rate.

**Table 15 Fixed Effects Parameter Estimates in SI Task**

*Fixed Effects Parameter Estimates in SI Task*

| Conditions | Effect  | Estimate | df   | t      | p      |
|------------|---------|----------|------|--------|--------|
| 1          | NF - NN | -20.66   | 63.0 | -17.92 | < .001 |
| 2          | DN - NN | -9.39    | 63.0 | -8.14  | < .001 |

Note: 95% Confidence Interval

**Table 16 Fixed Effects Parameter Estimates in ST Task***Fixed Effects Parameter Estimates in ST Task*

| Conditions | Effect  | Estimate | df   | t     | p      |
|------------|---------|----------|------|-------|--------|
| 1          | NF - NN | -2.00    | 63.0 | -2.80 | 0.007  |
| 2          | DN - NN | -3.08    | 63.0 | -5.31 | < .001 |

Note: 95% Confidence Interval

**5.5.3 The GSR Results**

Another supplemental indicator adapted in this study for measuring participants' cognitive load was the galvanic skin response (GSR). I calculated the mean GSR values of each participant under the NN, NF, DN and DF condition in the SI and ST tasks. Then, I performed the repeated measures ANOVA of GSR results. The findings show that there is no significant difference among the four different conditions. I also conducted the correlation analysis, and the results show that no significant relationship between GSR and accuracy score. In addition, no correlation was observed between pupil size and GSR based on the analysis conducted.

**Table 17 Correlation Result of GSR Data***Correlation Result of GSR Data*

|                        | 1         | 2     | 3 |
|------------------------|-----------|-------|---|
| 1. Mean Accuracy Score | —         |       |   |
| 2. Pupil Size Data     | -0.437*** | —     |   |
| 3. GSR Data            | 0.130     | 0.792 | — |

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

**5.5.4 The Working Memory Capacity Results**

In the lab experiment, I used two versions (visual and auditory) of the wordlist/word span test to replace the Corsi test to explore whether the WMC affects the interpreter's output (accuracy). The correlation results showed that both interpreters' auditory working memory span ( $r(20) = -0.038$ ,  $p = 0.617$ ) and visual working memory span ( $r(20) = -0.040$ ,  $p = 0.602$ ) has no relationship with their simultaneous interpreting or sight translation performance and vice versa.

**Table 18 Correlation Result of Working Memory Capacity***Correlation Result of Working Memory Capacity*

|                        | 1      | 2 | 3 |
|------------------------|--------|---|---|
| 1. Mean Accuracy Score | —      |   |   |
| 2. WMC (auditory span) | -0.038 | — |   |
| 3. WMC (visual span)   | 0.040  |   | — |

Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

## 5.6 Discussion

Following the results chapter, this discussion delves deeper into our exploration of the nuances in how variations in delivery rate and text difficulty influence trainees' cognitive load in both SI and ST tasks. In this chapter, I will discuss the findings, aligning them with established theoretical considerations and the results of prior research in the realms of SI and ST studies. Moreover, I will address the implications derived from the WMC and GSR results.

### 5.6.1 The Discussion of Research Questions

**Question 1: How does text difficulty and delivery rate impact on trainees' cognitive load during SI task?** Research using psycho-physiological methods – pupil size analysis – revealed that during SI tasks, trainees had a larger average pupil diameter in the Fast condition than in the Normal condition. However, text difficulty did not significantly affect the trainees' pupil size between the Difficult and Normal conditions. According to Tommola and Niemi's (1986, p. 175), "pupil dilation is related to the mental effort required by cognitive tasks, ... pupil size is correlated with the cognitive load experienced by a participant of an experiment". In general, pupils that are more dilated indicate higher cognitive load, while pupils that are less dilated indicate lower cognitive load, as supported by studies from Holmqvist et al. (2011) and Hvelplund (2014). Therefore, this result indicates that trainees experienced an increasing cognitive load as the speaker's delivery rate rises. However, this was not the case for an increase in text difficulty. The psychophysiological measures directly quantify the trainee interpreter's instant cognitive load change during the SI process. Building on this, performance

measurement allows us to evaluate how cognitive load change influences trainee interpreters' SI output. A deeper exploration of SI output can help us identify the cognitive load associated with production (Plevoets and Defrancq 2016). Transitioning to the aspect of performance, research using performance-based methodologies shows that the average holistic accuracy of the trainees in the Normal condition was higher than that in the Difficult condition. Moreover, their average holistic accuracy in the Normal condition was also higher than that in the Fast condition, with a notable difference. Subsequent statistical analysis reveals that a high delivery rate is predictive of poorer accuracy compared to the challenges posed by difficult text. Synthesizing these findings, one might conclude that, within the context of this study, the delivery rate exerts a more pronounced influence on trainees than text difficulty in the SI tasks. Supporting this conclusion, fourteen out of the twenty interviewed trainees conveyed that they encountered greater challenges under the high delivery rate condition during SI tasks.

Text difficulty as a challenge during SI has been proposed by many researchers (Selekovitch 1975; Gile 1995). Given the fundamental linguistic differences between Chinese and English in syntax, interpreters inevitably face huge challenges. Intriguingly, the present study revealed that during the SI process, source text difficulty did not elevate the trainees' cognitive load, as indicated by a lack of pupil size modulations. This contrasts with the self-assessment results from Experiment One. Several factors may account for these discrepancies. First, Experiment One relied on self-assessment methods, which inherently carry a degree of subjectivity, while lab study employed physiological measurements, offering a more objective perspective. Another consideration is that the source texts used in the two experiments were different, as were the participants. This could potentially account for the differing results between the two studies. Interestingly, although the texts used in Experiment Two were generally more difficult than those in Experiment One, this did not lead to a significant increase in the cognitive load of the interpreters. This could be attributed to the fact that the topics and styles of the lab texts resonated more with what trainees frequently encounter during their regular training. The findings underscore the importance of topic familiarity and background knowledge comprehension in SI and ST tasks. Furthermore, text difficulty variable tends to be rooted in "language-specificity related problems" (Gile, 2009, p. 194), encompassing issues like differential word perceptions, grammatical nuances, and syntactic variances. In the lab setting, when encountering these issues, I observed that trainees were able to employ certain strategies, like 'chunking', to effectively handle increased cognitive demands. Problems induced by text difficulty appear to be mitigated by familiarity with the topic and effective strategy application.

These methods help trainees to better distribute and manage their cognitive resources, ensuring they do not exceed their overall processing capacity. Most trainees in the interview said that they employed strategies to address the challenges posed by text difficulty.

On the other hand, under the condition of a high delivery rate, the situation becomes somewhat different. The result that a high delivery rate leads to higher cognitive load is in line with online study and other previous studies (Lee 1999; Seeber 2015). Barik (1973) and Gile (1995) both argued that a higher delivery rate might be a problem trigger in interpreting, and the effect of the fast delivery rate on SI output has been confirmed by some scholars (Gerver 1969; Pio 2003; Chang 2005; Liu 2004; Barghout 2015). However, most previous findings are based on quantitative research without qualitative analysis. The findings of this research may provide some empirical evidence, especially for studying English-Chinese language pairs.

In the present study, high delivery rates lead to an elevated cognitive load, as evidenced both by pupil size (reflecting the process) and accuracy (indicating the product). This increase might be attributed to the total requirements (including processing capacity and time) exceed the available cognitive capacity in the case of speech with a fast delivery rate, resulting in increased cognitive load on participants. Since it takes time to process words, the higher the speed, the more information needs to be processed per unit of time, which causes pressure on the Listening Effort; on the other hand, the same information also needs to be produced in the target language, and interpreters must not fall behind for afraid of memory saturation, so similar pressures apply to Production Effort (Gile 2021). In this case, high-delivery speech poses a great challenge to the balance of separate efforts. Some researchers (Altman 1994; Barik 1994; Gile 1995; Setton 1999) have regarded omissions in simultaneous interpreting either as errors or as technique interpreters might resort to under particularly challenging conditions, especially when faced with cognitive overload. In the present study, a higher frequency of word and segment omissions was observed in the Fast condition relative to the Difficult condition and Normal condition during SI tasks. Most of the terms omitted by the trainees were qualifying adjectives with high informative value. This finding is consistent with that of Pio (2003), who compared the effects of speech rate on student and professional interpreters. Another significant error that influences trainees' accuracy in the Fast condition are mainly due to the misinterpretation of numerals, which occurred more often than in Normal condition. Surprisingly, , twelve participants interpreted the number '*sixteen hundred*' to '*six hundred*', four participants interpreted this number to '*sixteen thousand*', three participants interpreted this number to '*six*

thousand’, and only five participants made a correct interpretation. Third-teen students replace the number ‘*millions*’ with a generic expression such as ‘很多’, ‘不少’ (many, a lot of). These results are in accord with Korpál’s (2016) findings. He compared student interpreters’ performance with professional interpreters’ performance for slow and fast delivery rate. His results indicated that although professional interpreters achieve lower error rates than students, in the slower delivery speed rate, nearly 30% of numbers in the interpreting task are either wrong or absent, while in the fast delivery rate task, the error rate jumps to 43%. Among the students, the error rate of numbers in the fast delivery rate task was 56%. Numeral has always been considered as a problem trigger in the interpreting task (Gile 1995; Braun and Clarici 1996; Mazza 2001; Korpál 2016). In this experiment, the increasing speed of speech rate makes numbers interpretation more challenging for interpreters.

In addition, to better understand the challenges imposed by a high delivery rate on information processing, I looked at what the trainees had to say. Their personal experiences give us a clear view of the real challenges they encounter:

- *Timing and Processing Constraints* – One trainee remarked, “Rapid speech impacts the time available to process information. Even when I comprehend the text, aligning the timing for my response becomes challenging.”
- *Lengthened Comprehension Process* – Another shared, “When the speaker delivers a high volume of words quickly, it lengthens my comprehension process, subsequently impacting my problem-solving abilities.”
- *Reduced Reaction Time* – Highlighting the direct consequences of fast speech, a trainee stated, “Increased delivery speed reduces my reaction time, giving me less opportunity to mentally process the tasks. This leads to the omission of more information.”

These first-hand accounts provide a microcosmic view of the overall challenges faced by trainees when dealing with high delivery rates, highlighting the effects on their cognitive processes. It is also worth noting that a high delivery rate imposes a significant mental strain on students, potentially leading to an increased cognitive load. To further explain this, I have selected responses from several trainees’ interviews:

- Trainee 1: I frustrated with high delivery rate in SI. I understood what the speaker was saying, but I could not keep up with the pace. This frustration progressively worsened my performance.

- Trainee 2: This seems like the first time I have directly felt the impact brought by the speed. I had previously anticipated some impact on my performance but did not expect it to be so significant. This unexpected challenge made me extremely nervous during the task, leading me to stop interpreting several times. There were several times I wanted to give up.
- Trainee 3: Once the speaker speeds her pace, I find it hard to keep up. This anxious feeling impacts my performance.
- Trainee 4: I can immediately sense a change in the speech's pace, and it adds a lot of pressure on me.

SI is a very complex and demanding task. A faster delivery rate necessitates rapid information processing. Trainees understand the content, but the pace does not allow them enough time to deliver the information accurately. Being unable to keep up with the pace may reduce their self-confidence. As they fall behind, the mounting pressure and realization of their performance lag may lead to heightened anxiety. High delivery may introduce an extraneous load that can be challenging for them. The change of speaker's speaking pace acts as an external factor, intensifying the extraneous cognitive strain. This suggests that in regular training sessions, trainees need to modulate their pace and optimize their cognitive resource allocation.

Another finding suggests that high delivery rate is a more significant predictor of interpreters' cognitive load than text difficulty. While the results from the statistical analysis indicate that a faster delivery rate has a greater influence on interpreters' cognitive load than text difficulty, it is important to note that the two variables, in essence, are not directly comparable. The intention is not to compare them in a competitive manner, but rather to highlight the distinct effects observed in the data. Furthermore, supporting the quantitative findings, qualitative interviews with the trainees showed that 15 out of 20 participants felt the impact of delivery rate was more pronounced than text difficulty in the context of SI tasks in the present study. One trainee highlighted the impact of delivery speed on comprehension, mentioning, "When the delivery is at a pace I can handle, I can make sense of difficult content through context. But as the speed increases, my understanding starts to reduce." Another trainee echoed this statement, emphasizing the limited effect of text difficulty: "I barely noticed changes in text difficulty, possibly due to my extensive training in handling difficult texts. In contrast, changes in delivery rate were immediately felt." The theme of the significant influence of high delivery rate

continued, with another trainee saying that “prediction strategy usually helps with difficult text, but it seems to falter when confronted with high delivery rate in SI”. Similarly, the difficulty in adjusting techniques between text difficulty and delivery rate was brought up by another trainee: “Some techniques I use for difficult texts does not work as well during high-speed delivery.” These facts seem to provide evidence for Gile’s (2021, p. 9) explanation of the Tighrope Hypothesis: “interpreters tend to work close enough to cognitive saturation for many EOIs to occur not because the interpreters’ insufficient knowledge of the working languages or the topics or insufficient technical skills but because attentional resources required to perform adequately were not available for a particular comprehension, memory storage or retrieval or production task at a time when they were needed.” This may explain why some students in the semi-structured interview said they comprehended the source text but could not ‘open their mouths’ under the fast delivery rate condition.

**Question 2: How does text difficulty and delivery rate impact on trainees’ cognitive load during ST task?** The accuracy results show that trainee achieve lower score in both the Difficult and Fast condition compared to their accuracy scores in the Normal condition. The change of text difficulty will impact ST products has been confirmed by some researchers (Chmiel and Mazur 2010; Li and Su 2019; Li 2019). The findings from the present study also align with previous research. Problem triggers such as the language-specific differences also significantly affect trainees’ accuracy in the ST task. Factors such as disparities in word frequency, sentence length, and readability influence the quality of the ST product. Beyond the challenges posed by the inherent difficulty of the source text, interference from the continuous availability of source information can also compromise ST task. With the source text consistently in view, trainees may find it challenging to concentrate on the overall meaning instead of focusing on specific words (Mikkelsen 1993). The “risk of linguistic interference” (Gile, 1997, p. 204) increases the complexity of the ST task. Interview from trainees also corroborates this. One participant noted, “In ST, if I get stuck on a word, my eyes remain fixated on it.” Another trainee mentioned, “I often revisit prepositions, which disrupts my ability to follow the sentence fluidly.” The format and layout of the text were also identified as critical; two trainees emphasized the importance of familiar text typesetting – “If the text layout is not what I am used to, it feels uncomfortable and may influence my output.” Additionally, sight translation lacks the auditory cues of a speaker’s intonation, emphasis, or pauses that assist with segmentation (Gile 2009). In the Difficult condition, the longer sentence length combined with the absence of these auditory signals can further complicate comprehension.

On the other hand, delivery rate/time pressure also significantly impacts trainees' accuracy. To my knowledge, no empirical study has focused on whether time pressure impacts on ST task. In the Fast condition, omissions are frequently accompanied by 'addition errors' that appear at the end of sentences in the target speech. This kind of addition matches with Barrik's (1994, p. 203) idea known as -- "closure additions", this kind of addition "accompanies rephrasing, omission or misinterpretation on the part of the T [source text] and which serves to give 'closure' to a sentence unit but does not add anything substantial to the sentence". I also noticed that most errors related to this category come with a longer eye-voice span. This might result from the prevalent strategy among trainees of reading an entire sentence at least once before starting sight translation. It is an effective way for them to grasp the key point and locate potentially critical information (Ma 2019). However, excessively extending the eye-voice span and relying on the strategy of reading the whole sentence before sight translating is not fully applicable to ST tasks under the high delivery rate. The reason is that according to our experiment setup, the sentences in the ST task disappeared from the screen in a limited time, so that strategy may not give them enough time to read the whole sentence before they start, leading to the loss of information. Under such time constraints, it is imperative for trainees to bolster their Coordination Effort to decrease errors. Furthermore, trainees also made a mistake in sight translating numerals. What is surprising is that in the fast delivery rate task, ten participants made mistakes when sight translating numerals. This outcome suggested that when performing time was limited, trainee interpreters may not be able to accurately sight translate the numerals even if they can see them in the ST task, because their information processing was affected by time pressure.

Turning our attention to another measurement, both text difficulty variable and delivery rate variable does not have a significant impact on the pupil size which is suggested to reflect the trainees' cognitive load during the ST process. The steady pupil size among varying delivery rates and text difficulties suggests that trainees might be efficiently distributing their cognitive resources during the ST tasks. This could possibly indicate that they have cultivated adaptive strategies to handle the accelerated pace or that the nature of ST provides some inherent flexibility in cognitive processing. A controlled and longer eye-voice span plus visually available text allows trainees to minimise the impact of the fast delivery rate and difficult text. Another possibility is that during the ST task, even with high delivery rates, might not overwhelm their brain, but it does require quicker decision-making. This rapid decision-making

could lead to more errors, reflected in the lower accuracy scores. Similarly, in the difficult condition during the ST task, while trainees are still processing information at a difficult level, they might be missing out on or glossing over finer details, which would again impact their accuracy without necessarily spiking their cognitive load. However, a decline in accuracy underscores a crucial point: while trainees are not perceptibly more burdened cognitively during the ST process, the rapidity and complexity inherent to high delivery rates and difficult texts will lead to more errors or lapses.

The results indicate that text difficulty is more likely to induce an increase in cognitive load in trainees. This observation is further substantiated by 13 trainees during the post-experiment interview, who stated that text difficulty affected them more than time pressure. For instance, one trainee mentioned that “It does not matter how fast the speaker’s speaking speed. If I know the content of the text, the delivery speed will not affect the processing process.” The predictability of this outcome is evident. In the ST task, the accuracy score difference between the normal delivery task and the fast delivery task is only 2 points, and only a small number of omissions and errors caused by the fast delivery rate were recorded. This phenomenon suggests that although there is a time constraint in the ST task, trainees can still regulate the eye-voice span to coordinate their output. However, the intensified interference stemming from increased text difficulty levels poses greater challenges to the trainees.

**Question 3: How does cognitive load differ between ST and SI tasks among interpreting trainees?** Research using psycho-physiological methods found that during SI tasks, trainees had a larger average pupil diameter compared to when they performed ST tasks. The research employing performance-based methodologies shows that the average of the trainees’ holistic accuracy in SI tasks was lower than in ST tasks. In contrast to ST, I also observed a higher occurrence of omissions in SI, which align with Agrifoglio’s (2004) findings. Piecing these observations together, it can be inferred that trainees’ average cognitive load during SI tasks was higher than that in ST tasks in the present study. This finding aligns with the online study. Corroborating this inference, fifteen of the twenty trainees in the interview expressed that SI is more challenging than ST. Although to my knowledge, no previous empirical studies have compared the cognitive load of trainee interpreters during the SI and ST task. SI has always been regarded as a high-load task (Gerver 1975; Darò 1989; Gile 1995; O’Brien 2006; Seeber 2011). The results due to the higher requirement for the ‘simultaneity’ of processing language under severe time constraints in the SI mode than in the ST mode, which makes SI task generate

more cognitive load. As we discussed in the online study, the nature of SI tasks demands a significant amount of predictive cognitive effort. The real-time nature of SI makes predictions more challenging compared to ST, where trainees can visually preview a sentence before translating. This viewpoint was echoed by three trainees during interviews. They noted that in ST, having the chance to visually assess a sentence in advance helps reduce the cognitive load. In contrast, SI tasks often bring inherent uncertainty, leaving them to guess the speaker's next term or chunk. The challenges intensify when there is a need to adjust word order, especially when interpreting prepositions, due to the marked differences between the Chinese and English language systems.

In addition, according to Gile (1995), eye-voice span in ST and ear-voice span in SI is evidence for the existence of the Effort Model and how the separate efforts are related to each other. A controlled eye-voice span in the ST task allows sight translators to reduce the burden of their Memory Effort and allocate more cognitive 'energy' to their Reading Effort and Production Effort. Chmiel et al. (2020) in their comparative study also confirmed that the time lag was shorter in SI than in ST. While the lab study places time restrictions on ST tasks, mirroring the duration seen in SI tasks with each sentence's screen display, participants retain the ability to adjust their pace. Trainee interpreters are assumed to reduce their cognitive load by regulating their EVS in ST task. The capacity for self-paced adjustment provides a considerable benefit, giving interpreters the chance to manage their rhythm throughout the task.

**Question 4: Does the impact of text difficulty and delivery rate on cognitive load differ between ST and SI tasks among interpreting trainees?** Our results indicate that both text difficulty and delivery rate significantly impact both SI and ST tasks, as reflected in the trainees' accuracy scores. However, when considering pupil size as an indicator, we found that a high delivery rate significantly affects the SI task but not the ST task. Additionally, text difficulty does not impact either the SI task or ST task. This observation is somewhat anticipated. Although the high delivery rate does influence ST, the discrepancy in accuracy scores between the Normal and Fast conditions for ST is a mere 2.78 points. In contrast, for the SI task, the difference is a substantial 22.01 points. Notably, the score difference for SI is nearly eight times that of ST. These results suggest that a well-coordinated effort could counter the increased cognitive load caused by problem triggers, such as a fast source text delivery rate. The extended eye-voice span plus visually available text potentially allows sight translators to minimise the impact of the fast delivery rate on their cognitive capacity. This introduces the possibility that

the simultaneous interpreting with text (SIT) mode could improve interpretation outcomes, particularly within the high delivery context of Chinese-English interpretation. However, a comprehensive comparison between the cognitive loads of SIT and SI is necessary to empirically substantiate this claim in the future.

Furthermore, intriguingly, during SI, the delivery rate has a more pronounced effect on trainees' cognitive load, whereas in ST, text difficulty exerts a greater influence on their cognitive load. In ST, the primary focus is on visual processing, with a significant emphasis on Reading Effort. When encountering a challenging text, this effort escalates, necessitating more cognitive resources for accurate comprehension and translation. Notably, the lack of auditory cues, which in other scenarios might assist with segmentation (Gile 2009), further amplifies the challenge in ST. On the other hand, SI mode emphasizing auditory processing and immediate recall. A fast delivery rate poses a direct challenge to the Listening Effort and Memory Effort, potentially broadening these cognitive resources to their limits and making it harder for the trainees to keep up. SI and ST are characterized by distinct cognitive models and information processing modalities. The unique challenges inherent to each method underscore the importance of tailored pedagogical approaches. Understanding these differences might provide empirical insights that enhance future training methodologies.

### **5.6.2 The Discussion of Working Memory Capacity Results**

In Experiment Two, we used word span test to measure trainees' working memory capacity (WMC). The results revealed that neither WMC auditory span nor WMC visual span correlated with accuracy scores. As mentioned before, working memory capacity variable served as the covariate variable, which is not the primary variable of interest. Although we recognize the relevance of the reading span and listening span for assessing WMC, our decision to utilize the corsi span and word span tests was driven by our main goal to consider WMC as a covariate rather than a main variable in our study. This could indicate that while WMC is generally crucial for cognitive tasks, it might not be a determinative factor affecting interpreting accuracy in this context, particularly from a performance-based perspective on cognitive load. This finding does not undermine the possible significance of working memory in general interpreting tasks. Rather, it implies that in evaluating individual discrepancies in SI and ST accuracy as seen in Experiment Two, variations in individual WMC might not be a pivotal consideration. However, this result still warrants reflection on the underlying reasons.

First, the lack of correlation between WMC and interpreting performance (accuracy score) could be attributed to the measure of memory span used in the present study is not well suitable for trainee interpreters. In recent decades, the best-known and most frequently used methods for measuring interpreters' or translators' WMC are probably the simple span paradigm (Corsi span, word span etc.) and the complex span paradigm (reading span, speaking span etc.). Corsi span is used for measuring spatial memory with minimal verbal mediation, and word span measurement more focus on verbal working memory with storage and maintenance function (Macnamara and Conway 2016). Oberauer (2004) defined three types of WMC measurements that load against a single WM factor: measuring the storage of briefly presented material parallel with interfering processing; measuring the construction of new structures and relationships; measuring some executive functioning. The span tests used by most interpreting scholars mainly focus on the first factor. However, simultaneous interpreting and sight translation both engage "complex and simultaneous storage and processing demands along with concurrent and continuous context-based restructuring" (Macnamara et al., 2011, p. 123). Most span tests used to measure the interpreter's WMC cannot fully capture the working memory capacity they need during ST and SI.

Another possible reason for the lack of correlation between WMC and performance is that the interpreter's WMC may play a minor role than other interpreting abilities in simultaneous and sight translation. This assumption was confirmed in Zhang's (2012) study, and he found that under the same conditions, interpreting skills have a greater impact on interpreting quality than WMC, especially when the task becomes more difficult, the interpreting expertise can help relieve the pressure of insufficient working memory resources. As Liu et al. (2004, p. 38) argued: "interpreter's expertise may rely much more on the acquisition of identifiable domain-specific skills than on general qualities such as a large working memory capacity". Furthermore, according to Gile (1995), the total available capacity and the available capacity for each separate effort are limited. Different from professional interpreters, it is hard for trainee interpreters to well coordinate all efforts. In order to avoid information overload, they may be more concerned with Listening Effort and Production Effort and adopt strategies related to these two efforts to solve the problem, at which point WMC was not the dominant element influencing their cognitive processes anymore. As Macnamara et al. (2011) mentioned, having a high working memory capacity does not benefit trainee interpreters if they cannot process

incoming information and make decisions either quickly or in line with the rate of incoming data.

### **5.6.3 The Application of Galvanic Skin Response in Lab Experiment**

In this study, the GSR value was used as an indicator of the trainee interpreter's cognitive load level. I collected the mean GSR value of the trainee interpreter under the four conditions in SI and ST task. The findings showed that GSR results did not correlate with interpreting accuracy scores and pupil size data. In addition, the difference in the GSR value of trainee interpreters in different tasks was also not significant. Regarding the interaction between variables and tasks, the high delivery rate condition has a slightly higher GSR score than normal delivery was observed in the SI task, but this difference was not significant.

Going back to the very beginning of this project, one of the purposes of the present study was to test the feasibility of GSR for empirical interpreting studies. The possible applications of GSR cover a wide range of academic and commercial research fields. Some scholars (Shi et al. 2007; Nourbakhsh et al. 2012; Ferreira et al. 2014; Johannessen et al. 2019) suggested that it might be used as an indicator of cognitive load. The GSR device sometimes is more comfortable to wear for interpreters than other neurology-method devices such as fMRI or EEG, especially when interpreters are asked to wear an eye tracker to allow researchers to track their eye movements. There is a lack of empirical studies on the applicability of GSR to measure the interpreter's cognitive load level during the SI and ST process. The outcomes of the present may provide some ideas for future exploration.

In the present study, GSR results associated with sweat gland measurements indicated that the reliability of this parameter as an indicator of cognitive load is limited. GSR is an ideal measure to track emotional arousal, but it is not able to reveal the emotional valence or the quality of the emotions (McCLeary 1950). Sweatiness will be affected by many variables such as the level of psychological arousal. However, arousal can be due to anger, fear, anxiety, excitement or nervousness, and all of these types of arousal can affect participants' activity in their autonomic nervous system (Wang et al. 2018). In the semi-structured interview of the present study, when the participants were asked how they felt about the experiment, twelve out of twenty-two participants thought that the whole experiment made them nervous, five students felt very stressed, and one student felt excited. All these emotions affect the changes in GSR

value. In this case, it is also possible that participants already at the elevated level of arousal due to the task they engaged in. Therefore, it is hard for us to see any changes on GSR value.

#### **5.6.4 Further Discussion of the Interview**

Some insights from the interviews also provided us with meaningful perspectives on SI and ST training. Fifteen participants reported experiencing emotional interference due to the high delivery rate in SI, while five participants noted similar interference in ST. This rapid pace appeared to induce feelings of anxiety and frustration during both the SI and ST procedures. Although there are very few studies on the emotional interference of the high delivery rate on ST or SI interpreters, Korpál (2016) has confirmed in his empirical study that a high delivery rate affects the psychological stress experienced by interpreters. Therefore, some participants believed that more training related to delivery rate could build their confidence and make them more positive in dealing with the difficulties brought by the speed in real interpreting tasks. This may remind trainers to consider more the role of delivery rate in SI and time pressure in ST during the daily training. However, one interesting finding was that participants who thought the text difficulty factor more influenced their performance than the delivery rate factor generally achieved higher accuracy scores than the other participants.

During interviews, three trainees highlighted a specific challenge they faced. One trainee remarked, “When I encounter fast speakers, my mind tends to drift, possibly due to difficulty in grasping the information. I believe this is tied to attention ability”. Another commented on the connection between attention and processing speed: “A tendency to lose focus can exacerbate difficulties when trying to keep up with high delivery speech”. The challenges of multitasking and the strain it places on cognitive resources have been discussed in numerous studies (Pashler 1994), and the phenomenon where the mind might ‘tune out’ is related to attentional lapses (Kahneman 1973). High delivery rates introduce a rapid inflow of information. Constantly processing this deluge can strain sustained attention, causing lapses as the interpreter may struggle to maintain consistent focus. Such observations underscore the potential value of incorporating attention training in interpreting training, especially when preparing for scenarios with high speech delivery in SI.

In addition, fifteen participants believe that ST training helps reduce their cognitive load during SI task. Students primarily found ST training beneficial for SI due to the perceived similarities

between the two modes. They highlighted factors such as the active problem-solving mechanism, the efficiency in practice time, and its suitability for beginners. Students emphasized that the similarities between the two modes, ST and SI, offer a unique advantage in interpreter training. One trainee remarked on the benefits of ST for improving expression, explaining, “Sight translation is helpful for expression. It is easier to understand a text by looking at it than by listening. Despite the distinct input mechanisms of listening and reading, their output methods are analogous, making their expression techniques similar. Therefore, mastering expression in ST directly improves SI proficiency.” Another participant drew a connection between visualizing content during SI and the act of ST, noting, “When I am doing SI, the terms and sentences I hear seem to be visualized, making it feel as though I am sight translating.” Several others emphasized the shared cognitive processes between ST and SI, with one highlighting, “Whether you are simultaneous interpreting or sight translating a text, the brain undergoes a similar processing to convert one language into another.” Another trainee spoke about the efficiency of ST, stating that “ST practice shortens reflection time because I am practicing while looking at the text.” Beginners particularly found value in ST as an introductory step, with one saying, “Starting with SI was challenging for me, so I chose ST. Practicing ST allows me to recall how a sentence should be outputted in SI from the memory of my own voice.” Lastly, the progression from ST to SI was considered an effective training methodology by a respondent who believed, “The interpretation training should be step by step, and transitioning from ST to SI is a suitable method. Since SI and ST have the same output style, ST helps in refining strategies like adjusting sentence structures, laying a foundation for SI.” These interview results all reflect the participants’ perception of the usefulness of ST training for SI training.

## Chapter 6 Conclusion

This research delved into the influences of delivery rate and text difficulty on the cognitive load experienced by trainee interpreters during simultaneous interpreting (SI) and sight translation (ST), utilizing both online and lab experimental approaches. This chapter will succinctly recap the research question, methods, and key findings, highlighting the significance — particularly the pedagogical implications for professional training in SI and ST. This section will also identify the study's limitations and offer insights into potential future directions for this line of inquiry.

### 6.1 Summary of the Research Findings

In revisiting the research questions of this study, I explored whether and how changes in text difficulty and delivery rate affect cognitive load in trainees during simultaneous interpreting (SI) and sight translation (ST) tasks. Additionally, this research aimed to determine if the impacts of text difficulty and delivery rate differ between SI and ST tasks. I further examined the distinctions in cognitive load between SI and ST to identify the specific differences. Regarding the research methods applied in this study, data collection encompassed both quantitative and qualitative methods, including self-assessments, pupillometry, galvanic skin response (GSR), holistic accuracy assessments, and retrospective interviews. The summary of research findings from the two experiments is presented in the following paragraphs.

In Experiment One, trainees faced a greater cognitive load in SI compared to ST. In SI, both delivery rate and text difficulty influenced perceived cognitive load and accuracy scores. Across both methods, the impact of delivery rate was more pronounced than text difficulty. In ST, both variables affected perceived cognitive load and accuracy scores, but no significant difference was found between the effects of delivery rate and text difficulty on perceived cognitive load. Additionally, in ST, text difficulty had a stronger influence on accuracy than delivery rate. There is no observed correlation between WMC and either accuracy or perceived cognitive load.

In Experiment Two, SI posed greater cognitive challenges than ST. In the realm of SI, while delivery rate influenced pupil size as an indicator of cognitive load, both variables (delivery

rate and text difficulty) affected accuracy scores, with the former having a greater influence. For ST tasks, neither delivery rate nor text difficulty markedly affected pupil size, though both had significant effects on accuracy scores, underscoring the larger impact of text difficulty in this mode. Both WMC auditory span and WMC visual span exhibited no correlations with accuracy scores. Lastly, no significant differences were observed in GSR across various conditions in both SI and ST tasks, and no correlation was identified between GSR and either pupil size or accuracy scores.

In summary, trainees experience a higher cognitive load in sight translation compared to simultaneous interpreting. Regarding two crucial input variables, delivery rate changes have a more pronounced effect on cognitive load in SI, whereas in ST, the variations in text difficulty play a more dominant role in influencing cognitive load. The differences in cognitive load across these interpreting modes might be attributed to their distinct cognitive demands, for instance, simultaneous interpreting leans more towards auditory processing, while sight translation hinges on visual comprehension. Furthermore, the results from this study raise questions regarding the suitability of GSR as a tool for accurately measuring cognitive load in interpreting and sight translation.

## **6.2 Pedagogical Implications on SI and ST Training**

The findings from this study reveal key pedagogical implications for the professional training of simultaneous interpreting and sight translation, warranting a separate section to explore their importance and applications. In SI, the pronounced impact of delivery rate suggests that cognitive problem triggers (Gile, 2009, p. 192) have a more significant effect on the trainee interpreter's performance than language-specificity related problems (Gile, 2009, p. 194), ultimately affecting their interpreting accuracy. This observation underscores the need for trainees to improve their cognitive coordination skills to process information more effectively under conditions of high delivery rates. To address this, regular training sessions should aim to guide trainees in adjusting their pace and optimizing the allocation of cognitive resources. This guidance is vital for developing strategies that effectively manage cognitive resources, which is crucial for reducing cognitive load in high delivery rate scenarios during SI. Additionally, the high delivery rate in SI leads to emotional interference, as the swift pace induces feelings of anxiety and frustration among trainees during SI procedures. This may highlight the

importance for trainers to pay closer attention to the role of delivery rate in SI during regular training. Increased focus on training related to delivery rate can build confidence and foster a more positive attitude among trainees towards handling the challenges presented by speed in actual interpreting tasks.

In ST, text difficulty exerts a more pronounced impact on trainees. The sustained presence of the source language increases the likelihood of interference between the two languages, potentially impacting ST performance (Agrifoglio 2004; Gile 2009; Shreve 2010). This scenario presents significant challenges for trainees, necessitating enhanced training and the adoption of anti-interference strategies to mitigate visual interference. While the delivery rate is less influential during ST tasks, it still affects their accuracy, particularly in the translation of numerals. In conditions of high delivery rates, omissions often coincide with ‘addition errors’ at sentence endings in the target speech. This phenomenon supports Barrik’s (1994, p. 203) concept of “closure additions”. Errors in this category are typically associated with a longer eye-voice span, likely due to trainees’ common practice of reading an entire sentence before beginning the sight translation. However, overly extending the eye-voice span and relying solely on reading the sentence in full before translating proves less effective in ST tasks with high delivery rates. Thus, ST training sessions should guide trainees to adjust their eye-voice span, especially when facing time-limited tasks.

Furthermore, during SI, the delivery rate significantly impacts trainees’ cognitive load, whereas text difficulty plays a more crucial role in ST. This indicates that these two modalities engage different cognitive processes and information handling, necessitating distinct pedagogical strategies. While many students, as revealed by interviews, acknowledge the benefit of sight translation exercises in interpreting training, and numerous higher education institutions have incorporated sight translation into their interpreting modules, it is vital to recognise and address the unique challenges posed by SI and ST. This distinction underscores the need for differentiated training approaches in curriculum development, ensuring that each method’s specific demands are adequately met.

Moreover, two extensions from the research findings related to interpreting pedagogy involve the application of the simultaneous interpreting with text (SIT) mode and the potential value of attention training. According to the present research, the extended eye-voice span combined with visually available text potentially enables sight translators to minimise the cognitive

impact of rapid delivery rates. This introduces the possibility that the SIT mode could enhance interpreting performance in high-delivery contexts of Chinese-English interpretation. Additionally, some interviews highlighted that high delivery rates introduce a rapid influx of information. Constantly processing this information can strain sustained attention, causing lapses as the interpreter may struggle to maintain consistent focus. These observations underscore the potential benefits of integrating attention training into interpreting education, particularly in preparing for high-speed speech delivery scenarios in SI.

### **6.3 Significances and Limitations**

This study offers some contributions to the field of Interpreting and Translation Studies, shedding light on key areas.

Cognitive load and the allocation of cognitive resources are drawing heightened attention in Translation and Interpreting Studies. This research undertakes a comparative analysis of two distinct interpreting modes, both involving different cognitive mechanisms: sight translation and simultaneous interpreting. It offers valuable insights that can be instrumental for more informed comparative studies in the domain. By investigating the interplay between source input variables and cognitive load across varied delivery rate and text difficulty conditions, the findings not only augment the existing scholarly discourse but also provide empirical evidence for verifying the tight rope theory.

Furthermore, a feature of the present study lies in its methodological design. This study stands out as one of the early endeavours in Translation and Interpreting Studies that employs a mixed-methods framework to investigate the cognitive load dynamics of English to Chinese SI and ST tasks. Notably, our approach is distinguished by its comprehensive measurement of cognitive load, utilizing four distinct methods, thereby providing depth to our understanding of cognitive load. The combined use of online and lab experiments has not only expanded the methodological horizons but has also offered a reference point for studies conducted under special circumstances. Additionally, with physiological metrics gaining traction in Interpreting and Translation Studies, our incorporation of the GSR indicator offers an invaluable reference for subsequent research. This study also provides insights into the assessment of working memory capacity, advancing our understanding of its role and implications.

Pedagogically, this study offers strategies to trainees grappling with challenges related to high delivery rate and difficult text in SI and ST task. In addition, training models should align with the distinct challenges posed by different modes. For sight translation, it is suggested that the focus may need to be on comprehending text difficulty, given its significant impact on trainees' cognitive performance. Conversely, in simultaneous interpreting, training may need to prioritise equipping students with strategies to handle changes in delivery rate, as these variations have been demonstrated to exert a greater influence on their output.

Lastly, this research, which integrates principles from both Interpreting and Translation Studies and Cognitive Psychology, reflects the growing trend of interdisciplinary exploration. It is clear that interdisciplinary efforts are having a major impact on academic research. Such approaches use the strengths from various disciplines, leading to a more complete understanding and encouraging collaboration between researchers. This study not only helps us better understand the topic of cognitive load but also offers insights into bilingual task-oriented cognitive research.

While the present study offers above insights, there are inherent limitations that need acknowledgment.

One of the contentious issues in an empirical study is the difference between a controlled experiment and naturalistic observation. The advantage of the experimental design is that researchers can draw conclusions about the causal relationship between variables by testing the influence of an independent variable on a dependent variable. However, experiments are often criticised for their lack of relatability to real-world experience. In studies on interpreters requiring experimental manipulation, researchers must design the speeches to be interpreted, which may decrease the ecological validity (Korpala 2016). For example, in the ST task in this lab study, the experimental texts presentation differed from the texts in some real ST tasks. In the present experiment, each sentence in the ST task was individually presented in consecutive slides to reduce the influence of visual interference from the source text on trainee interpreters. However, such a setting has reduced the ST task's difficulty and may raise participants' accuracy scores. I also acknowledge that conveying sight translation texts at the speed of simultaneous interpreting is an ambitious approach that requires further refinement in the future.

Moreover, it is not easy to control all the extra variables in an experiment involving human participants because each participant has different experiences that can influence their behaviour. The participants know they are being experimented on and may behave differently than they would in real life, increasing the risk of demand characteristics and social desirability bias. Additionally, due to the challenges posed by the pandemic, the sample size for our in-person experiments was not as extensive as we would have preferred, which might have some impact on the interpretation of our results.

#### **6.4 Future Avenues of the Research**

Having discussed the potential limitations of the study, I will proceed to areas for further research.

Firstly, a possibility for extending WMC research is to consider different measuring methods. While we acknowledge the potential value of the reading span and listening span in gauging WMC, our choice of the corsi span and word span tests was informed by our primary intent to focus on WMC as a covariate, rather than a main variable of interest. Throughout the course of this research, the varied methodologies employed by others in the field to measure WMC have highlighted the intricacies in this domain. This leads to a question: whether we can develop various methods to capture better the processing and storage capacities needed during SI and ST, rather than traditional or common memory capacity approaches. For example, we can consider applying a test involving strategic chunking to the interpreter's WMC measurement. Chunking as a strategy in simultaneous interpreting and sight translation refers to the process whereby interpreters segment a source sentence into smaller fragments and encode them sequentially (Seeber 2011; Kader and Seubert 2014; Ahrens 2017). Chunking is driven by the unit of meaning, and a meaning unit can be as short as a word which depending on various factors, such as the interpreter's WMC (Miller 1956). A chunk integrates several segments expanding the interpreter's storage capacity and releasing more cognitive resources for other operations (Gile 2008; Nobre et al. 2013). The phenomenon of 'chunking' as a memory mechanism related to WMC measurement has been used in Cognitive Psychology Studies (Jefferies et al. 2004; Germano et al. 2008; Kapikian and Briscoe 2012). However, the application of the concept of chunking to the measurement of WMC has not been widely concerned in Interpreting Studies. How to develop a test suitable for measuring the interpreter's

WMC, such as a set of binding verbal and visuospatial material strategic chunking or clustering of information) is a worthwhile exploration in the future.

Second, in the context of high delivery English to Chinese interpretation, the simultaneous interpreting with text (SIT) mode may offer improved interpretation outcomes. As mentioned before, one of the findings from this study suggests there is potential in this direction. Given this, a valuable direction for future research would be a comprehensive comparison between the SIT and SI modes to evaluate the validity of this proposition. Besides, another promising avenue related to high delivery rate is the exploration of the relationship between attention and high speech rates or time pressure, as referred to in the main text. The intricate interplay between an individual's attentional capacity and the demands of rapid information processing under high delivery rates can have profound implications for the interpreter's performance. A deeper understanding of attention mechanisms in SI and ST may offer insights into training strategies and adaptive techniques for interpreters in real-world scenarios.

Furthermore, this study investigates trainees' cognitive load from an English to Chinese perspective. However, directionality is indeed a topic worthy of exploration in the Interpreting and Translation Studies. Therefore, future research might consider examining cognitive load in distinct modes from a Chinese to English perspective. In addition, this research focuses solely on trainee interpreters. Future studies could compare the cognitive load between trainee interpreters and professional interpreters across different modes, which would greatly aid in providing more comprehensive strategies.

Lastly, further application of physiological metrics like GSR in the realms of interpreting and translating remains an ongoing endeavour. This involves more comprehensive testing of measurement methods and potentially giving more attention to the emotional arousal experienced during the interpreting and translating process.

## References

1. Adhikari, P., Zele, A. and Feigl, B. (2015) The Post-Illumination Pupil Response (PIPR). *Investigative Ophthalmology and Visual Science*, 56 (6), pp. 3838-49.
2. Agrifoglio, M. (2004) Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, 6 (1), pp. 43–67.
3. Ahrens, B. (2017). Interpretation and Cognition, In Schwieter, J. W., Ferreira, A., & Wiley, J. (Eds.). *The handbook of translation and cognition*. New York: Wiley- Blackwell.445-460.
4. Ais, A. and Becerra, O. Quality Criteria In: F. Pöchhacker, ed. *Routledge Encyclopedia Of Interpreting Studies*. , pp. 337-338.
5. Albir, A. and Alves, F. (2008) Translation as a cognitive activity. In: J. Munday, ed. *The Routledge Companion to Translation Studies*. 1 ed. London and New York: Routledge, pp. 54-73.
6. Albl-Mikasa, M, Tiselius and E (2021) *The Routledge Handbook of Conference Interpreting*. Routledge.
7. Alexandridis, E. (1985) *Pupil Size*. New York: Springer.
8. Altman, J. (1994) Error analysis in the teaching of simultaneous interpreting: A pilot study. In S. Lambert & B. Moser-Mercer (eds) *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam/Philaelphia: John Benjamins, 25–38.
9. Alves, F. and Albir, A.H. (2017) Evolution, Challenges, and Perspectives for Research on Cognitive Aspects of Translation. In: *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, IncApr 19, pp. 535-554.
10. Aron, A., Aron, E. and Coups, E. (2014) *Statistics for Psychology*. 6 ed. Edinburgh: Pearson Education Limited.
11. Babcock, L. and Vallesi, A. (2017) Are simultaneous interpreters expert bilinguals, unique bilinguals, or both? *Bilingualism: Language and Cognition*, 20 (2), pp. 403-417.
12. Backs, W. and Seljos, A. (1994) Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *International Journal of Psychophysiology*, 16 (1), pp. 57–68.
13. Baddeley, A., D. and Hitch, G. (1974) Working memory. In: G. Bower H., ed. *The psychology of learning and motivation*. New York: Academic Press, pp. 47-89.
14. Baigorri and J (2014) *From Paris to Nuremberg*. Benjamins Translation Library.

15. Baker, M. (1998) *In Other Words: A Course Book on Translation*. London and New York: Routledge.
16. Balota, D. and Chumbley, J. (1984) Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology*, 10 (3), pp. 340–357.
17. Barghout, A., Rosendo, L.R. and García, M.V. (2015) The influence of speed on omissions in simultaneous interpretation: An experimental study. *Babel*, 61 (3), pp. 305-334.
18. Barik, H. (1973) Simultaneous interpretation: Temporal and quantitative data. *Language and Speech*, 16 (3), pp. 237-270.
19. Barrouillet, P. and Camos, V. (2012) As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, 21 (6), pp. 413-419.
20. Bates, D. and DebRoy, S. (2004) Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, 91 (1), pp. 1-17.
21. Bayliss, D.M., Jarrold, C., Baddeley, A.D. and Gunn, D.M. (2005) The relationship between short-term memory and working memory: Complex span made simple? *Memory (Hove)*, 13 (3-4), pp. 414-421.
22. Beatty, J. (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91 (2), pp. 276-292.
23. Berch, D.B., Krikorian, R. and Huha, E.M. (1998) The Corsi Block-Tapping Task: Methodological and Theoretical Considerations. *Brain and Cognition*, 38 (3), pp. 317-338.
24. Bertelson, P., Mousty, P. and Monique, R. (1992) The time course of braille word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18 (2), pp. 284–297.
25. Binda, P. and Murray, S. (2015) Keeping a large-pupilled eye on high-level visual processing. *Trends Cogn Sci*, 19, pp. 1–3.
26. BIOPAC Systems, I. (2014) *Biopac Student lab Manual*.
27. Boller, F., Keefe, N., C. and Zoccolotti, P. (1989) Luigi Galvani, body electricity, and the 'galvanic skin response'. *Neurology*, 39 (6), pp. 868-870.
28. Borsboom, D. (2005) *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.

29. Boucsein, W. (2012) *Electrodermal activity*. 2. ed. ed. New York
30. Bourne, V. (2017) *Strating out in methods and statistics for psychology*. Oxford: Oxford University Press.
31. Braun, S. and Clarici, A. (1996) Inaccuracy for Numerals in Simultaneous Interpretation: Neurolinguistic and Neuropsychological Perspectives. *The Interpreters' Newsletter*, 7, pp. 85–102.
32. Broadbent, D., E (1958) *Perception and Communication*. London: Pergamon Press.
33. Bryman, A. (2016) *Social Research Methods*. 5 ed. Oxford: Oxford University Press.
34. Cacioppo, J.T., Berntson, G.G. and Nusbaum, H.C. (2008) Neuroimaging as a New Tool in the Toolbox of Psychological Science. *Current Directions in Psychological Science*, 17 (2), pp. 62-67.
35. Carpenter, P. and McDonald, L. (1981) Syntactic and semantic influences in parsing ambiguous phrases. Pittsburgh: Carnegie-Mellon University.
36. Čeňková, I. (2015) Sight Interpeting/Translation. in Pöchhacker, F. (ed). Routledge encyclopedia of interpreting studies. Routledge, 374-375
37. Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied cognitive psychology*, 10(2), 151-170.
38. Chang, C. (2009) Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality. PhD, Department of Humanities Imperial College London University of London.
39. Chen, S. (2017) The construct of cognitive load in interpreting and its measurement. *Perspectives: Studies in Translation Theory and Practice*, 25 (4), pp. 640-657.
40. Chernov, G., V. (2004) *Inference and anticipation in simultaneous interpreting: A probability-prediction model*. Amsterdam: John Benjamins.
41. Cheung, A. K. (2013). Non-native accents and simultaneous interpreting quality perceptions. *Interpreting*, 15(1), 25-47.
42. Cherry, K. (2022) *Within Subjects Design and Between Subjects Design*.
43. Chincotta, D. and Underwood, G. (2006) Simultaneous interpreters and the effect of concurrent articulation on immediate memory. *Interpreting*, 3, pp. 1-20.
44. Chmiel, A. and Mazur, I. (2010) Eye tracking sight translation performed by trainee interpreters. In: C. Way et al., eds. *Tracks and Treks in Translation Studies*. Amsterdam / Philadelphia: John Benjamins, pp. 189-205.

45. Chmiel, A. (2018) In search of the working memory advantage in conference interpreting – Training, experience and task effects. *International Journal of Bilingualism*, 22 (3), pp. 371–384.
46. Chmiel, P. (2020) The eye or the ear? *International Journal of Research and Practice in Interpreting*, 22 (2), pp. 187-210.
47. Christoffels, I., Groot, A. and Waldorp, L. (2003) Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition*, 6 (3), pp. 201-211.
48. Christoffels, I., de Groot, A. and Kroll, J. (2006) Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54 (3), pp. 324-345.
49. Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O. and Engle, R.W. (2005) Working Memory Span Tasks: A Methodological Review and User's Guide. *Psychonomic Bulletin & Review*, 12 (5), pp. 769-786.
50. Corsi, P.M. (1972) *Human memory and the medial temporal region of the brain*. PhD, McGill University. [Accessed: 20th August 2021].
51. Cowan, N. (1995) Attention and memory: An integrated framework. In: *Oxford Psychology Series 26*. New York: Oxford University Press.
52. Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., Barral, O., Kosunen, I., Valtonen, T., Huutilainen, M., Ravaja, N. and Jacucci, G. (2016) The Psychophysiology Primer. *Foundations and Trends in Human-Computer Interaction*, 9 (3-4), pp. 151-308.
53. Dam, H. V. (2001). On the option between form-based and meaning-based interpreting: The effect of source text difficulty on lexical target text form in simultaneous interpreting. *The Interpreters' Newsletter*, 11, 27-55.
54. Darò, V. (1994) Non-linguistic factors influencing simultaneous interpretation. In: S. Lambert and B. Moser-Mercer, eds. *Bridging the gap: empirical research in simultaneous interpretation*. Philadelphia: John Benjamins, pp. 249-272.
55. Dehue, N. and van de Leemput, C. (2014) What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, 5, pp. 1099.
56. Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language teaching*, 42(4), 476-490.

57. Dillinger, M. (1994) Comprehension during interpreting: What do interpreters know that bilinguals don't? In: S. Lambert and B. Moser-Mercer, eds. *Bridging the Gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, pp. 155-189.
58. Dragsted, B. and Hansen, I. (2009) Exploring Translation and Interpreting Hybrids. The Case of Sight Translation. *Meta*, 54 (3), pp. 588–604.
59. Ebitz, R. and Moore, T. (2017) Selective modulation of the pupil light reflex by microstimulation of prefrontal cortex. *J Neurosci*, 37, pp. 5008–5018.
60. Ebitz, R. and Moore, T. (2019) Both a Gauge and a Filter: Cognitive Modulations of Pupil Size. *Frontiers in Neurology*, 9, pp. 97-110.
61. Engle, R. W. and Kane, M. J. (2004) Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 44, pp. 145–199.
62. Ericsson, K. A., & Simon, H. A. (1984/1991). Protocol analysis: Verbal reports as data. the MIT Press.
63. European Commission *Simultaneous Interpreting*. Available: [https://knowledge-centre-interpretation.education.ec.europa.eu/en/conference-interpreting/simultaneous-interpreting#:~:text=Simultaneous%20interpreting%20is%20a%20mode,same%20time%20\(or%20simultaneously\).](https://knowledge-centre-interpretation.education.ec.europa.eu/en/conference-interpreting/simultaneous-interpreting#:~:text=Simultaneous%20interpreting%20is%20a%20mode,same%20time%20(or%20simultaneously).)
64. Falbo, C. (2015) Error Analysis. in Pöchhacker, F. (ed). Routledge encyclopedia of interpreting studies. Routledge, 143-144
65. Féal, K.D. (1990) Some Thoughts on the Evaluation of Simultaneous Interpretation. In: *Interpreting-Yesterday*. State U of New York, pp. 154.
66. Ferreira, F., Ferreira, D., Kim, S., Siirtola, P., Röning, J., Forlizzi, J.F. and Dey, A.K. (2014) Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults. *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. Orlando, FL, USA, 9-12 Dec., pp. 39-48.
67. Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 32 (3), pp. 221–233.
68. G. Seeber, K. and Zelger, C. (2007) Betrayal – Vice or Virtue? An Ethical Perspective on Accuracy in Simultaneous Interpreting. *Meta (Montréal)*, 52 (2), pp. 290-298.
69. Gaiba, F. (1998) *The Origins of Simultaneous Interpretation: The Nuremberg Trial*. University of Ottawa Press.
70. Galy, E., Cariou, M. and Mélan, C. (2012) What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83 (3), pp. 269-275.

71. Carl, M. (2009). Triangulating product and process data: quantifying alignment units with keystroke data. In Mees, I. M., Göpferich, S., & Alves, F. (eds). *Methodology, technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (Copenhagen Studies in Language 38). Samfundslitteratur, 225-247.
72. Garzone, G. (2003) Reliability of Quality Criteria Evaluation in Survey Research. In: *La evaluación de la calidad en interpretación: Investigación*. Interlingua Institute, pp. xi.
73. Gelman, A. and Hill, J. (2007) *Data analysis using regression and multilevel, hierarchical models*. 1. publ. ed. Cambridge [u.a.]: Cambridge Univ. Press.
74. Gerjets, P., Scheiter, K. and Catrambone, R. (2004) Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science*, 32 (1/2), pp. 33-58.
75. Germano, C., Kinsella, G., Storey, E., Ben, O. and David, A. (2008) The episodic buffer and learning in early Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 30 (6), pp. 627-638.
76. Gerver, D. (1969/2002). The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. Republished in: Pöchhacker, F. & Shlesinger, M. (eds.). *The Interpreting Studies Reader*. London and New York: Routledge. 52-66.
77. Gerver, D. (1975) A psychological approach to simultaneous interpretation. *Meta: Translators' Journal*, 20 (2), pp. 119-128.
78. Ghobadi, M., Golnaz, M. and Bahareh, N. (2017) A Study of the Effects of Time Pressure on Translation Quantity and Quality. *International Journal of Comparative Literature and Translation Studies*, 5 (2), pp. 7-13.
79. Gieshoff, A. (2021) Does it help to see the speaker's lip movements? An investigation of cognitive load and mental effort in simultaneous interpreting. *Translation, Cognition & Behavior*, 4 (1), pp. 1-25.
80. Gile, D. (2008) Local Cognitive Load in Simultaneous Interpreting and its Implications for Empirical Research. *Forum*, 6 (2), pp. 59-77.
81. Gile, D. (2021) The Effort Models of interpreting as a didactic construct. *New Frontiers in Translation Studies*, , pp. 139-160.
82. Gile, D. (1995) Fidelity Assessment in Consecutive Interpretation: An Experiment. *Target : International Journal of Translation Studies*, 7 (1), pp. 151-164.

83. Gile, D. (2003) Quality Assessment in Conference Interpreting: Methodological Issues. In: *La evaluación de la calidad en interpretación: Investigación*. Interlingua Institute, pp. xiii.
84. Gile, D. (2009) *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins Publishing Company.
85. Gile, D. (2016) *The Effort Models - Clarifications and update*.
86. Goff, D.A., Pratt, C. and Ong, B. (2005) The Relations between Children's Reading Comprehension, Working Memory, Language Skills and Components of Reading Decoding in a Normal Sample. *Reading & Writing*, 18 (7-9), pp. 583-616.
87. Goldberg, J. and Wichansky, A. (2003) Eye Tracking in Usability Evaluation: A Practitioner's Guide. In: J. Hyönä et al., ed. *The Mind's Eye Cognitive and Applied Aspects of Eye Movement Research*. pp. 493-516.
88. Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech. Academic Press, London & New York.
89. Golla, F., L. (1921) Croonian Lectures on The Objective Study of Neurosis. *The Lancet*, 198 (5110), pp. 265-270.
90. Granholm, E. and Steinhauer, S.R. (2004) Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, 52 (1), pp. 1-6.
91. Gumul, E. (2021) Explicitation and cognitive load in simultaneous interpreting Product- and process-oriented analysis of trainee interpreters' outputs. *Interpreting*, 23 (1), pp. 45-75.
92. Guo, Y. and Yang, L. (2019) *实验心理学*. 2 ed. Beijing: People's Education Press.
93. Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: *Advances in Psychology*. The Netherlands: Elsevier Science & Technology, pp. 139-183.
94. Hale, S. and Napier, J. (2013) *Research Methods in Interpreting: A Practical Resource*. 1 ed. Bloomsbury Publishing.
95. Healey, J, F. (2015) *The essentials of statistics: a tool for social research*. 4 ed. Cengage Learning, Inc.
96. Hess, H. and Polt, M. (1960) Pupil size as related to interest value of visual stimuli. *Science*, 132 (3423), pp. 349-350.
97. Hild, A. (2011). Effects of linguistic complexity on expert processing during simultaneous interpreting. In Alvstad, C., Hild, A., & Tiselius, E. (eds.). *Methods and strategies of process research*:

Integrative approaches in Translation Studies (Vol. 94). Amsterdam/Philadelphia: John Benjamins Publishing, 249-267.

98. Hoeks, B. and Levelt, W. (1993) Abstract only.
99. Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011) Eye Tracking: A Comprehensive Guide To Methods And Measures. In: J. Weijer, ed. *Eye tracking: a comprehensive guide to methods and measures*. Oxford: Oxford University Press.
100. Howes, D. and Solomon, R. (1951) Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41 (6), pp. 401-410.
101. Hvelplund, K. T. (2011). Allocation of cognitive resources in translation: An eye-tracking and key-logging study. *Samfundslitteratur*.
102. Hvelplund, K. (2014) Eye tracking and the translation process: reflections on the analysis and interpretation of eye tracking data. *MonTI Special Issue – Minding Translation*, , pp. 201-233.
103. Hyönä, Jukka, Tömmola, Jorma & Alaja, Anna-Mari (1995) Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology Section A* 48 (3), 598–612.
104. Injoque-Ricle, I., Barreyro, J., Formoso, J. and Jaichenco, V. (2015) Expertise, Working Memory and Articulatory Suppression Effect: Their Relation with Simultaneous Interpreting Performance. *Advances in Cognitive Psychology*, 11 (2), pp. 56-63.
105. Jakobsen, A. (2017) Translation Process Research. In: J. Schwieter and A. Ferreira, eds. *The Handbook of Translation and Cognition*. 1 ed. John Wiley & Sons, pp. 21-49.
106. Jefferies, E., Lambon, R., Matthew, A. and Baddeley, D. (2004) Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, 51 (4), pp. 623-643.
107. Jensen, K. T. (2009). Indicators of text complexity. In Göpferich, S., Jakobsen, A. L., & Mees, I. M. (eds.). *Behind the mind: Methods, models and results in translation process research* (Vol. 37). *Samfundslitteratur*, 61-80.
108. Jereščenková, A. (2014) *Interferenzen beim Vom-Blatt-Dolmetschen und beim Simultandolmetschen*. Master, University of Vienna.
109. Jescheniak, J. and Levelt, W. (1994) Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology*, 20 (4), pp. 824–843.

110. Johannessen, E., Szulewski, A., Radulovic, N., White, M., Braund, H., Howes, D., Rodenburg, D. and Davies, C. (2019) Psychophysiological measures of cognitive load in physician team leaders during trauma resuscitation. *Computers in Human Behavior*, 111.
111. Jones, D., Farrand, P., Stuart, G. and Morris, N. (1995) Functional Equivalence of Verbal and Spatial Information in Serial Short-Term Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 21 (4), pp. 1008-1018.
112. Jones, R. (2008) *会议口译解析*. Shanghai: Shanghai Foreign Languages Education Press.
113. Jones, R. (1998) *Conference interpreting explained*. 1. publ. ed. Manchester, UK: St. Jerome Publ.
114. Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329.
115. Kader, S. and Seubert, S. (2014) Antipation, segmentation...stalling? How to teach interpreting strategies. In: D. Andres and B. M, eds. *To Know How to Suggest .... Approaches to Teaching Conference Interpreting*. Frank & Timme.
116. Kahane, E. (2015) Thoughts on the quality of interpretation. *Aiic*, .
117. Kahneman, D. (1973) *Attention and effort*. Englewood Cliffs.
118. Kahneman, D. and Beatty, J. (1966) Pupil diameter and load on memory. *Science*, 154 (3756), pp. 1583-1585.
119. Kantowitz, B., Roediger III, H. and Elmes, D. (2008) *Experimental Psychology*. 9 ed. Belmont: Cengage Learning.
120. Kapikian, A. and Briscoe, J. (2012) Semantic binding, not attentional control, generates coherent global structure in children's narrative memory. *Journal of Cognitive Psychology*, 24 (6), pp. 751-764.
121. Kelbsch, C., Strasser, T., Chen, Y., Feigl, B., Gamlin, P.D., Kardon, R.H., Peters, T., Roecklein, K.A., Steinhauer, S.R., Szabadi, E., Zele, A.J., Wilhelm, H. and Wilhelm, B. (2019) Standards in Pupillography. *Frontiers in Neurology*, 10, pp. 129.
122. Kessels, R., P., Zandvoort, M, J, E., Postma, A., Kappelle, L., J. and Hand, E, H, F. (2000) The Corsi Block-Tapping Task: standardization and normative data. *Applied Neuropsychology*, 7 (4), pp. 252-258.
123. Khawaji, A., Zhou, J., Chen, F. and Marus, N. (2015) Using Galvanic Skin Response (GSR) to Measure Trust and Cognitive Load in the Text-Chat Environment. *Proceedings of the 33rd Annual*

- ACM Conference Extended Abstracts on Human Factors in Computing Systems. New York, United State, April 18 - 23. Association for Computing Machinery, pp. 1989–1994.
124. Kiraly, D. (2000) *A Social Constructivist Approach to Translator Education*. Taylor and Francis. Available: <https://www.taylorfrancis.com/books/9781317641407>.
  125. Klare, G. R. (1984) Readability. In: Pearson, P. D. & Barr, R. (eds) *Handbook of Reading Research*. New York: Longman. 681–744.
  126. Klepsch, M., Schmitz, F. and Seufert, T. (2017) Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, 8, pp. 1997.
  127. Köpke, B. and Nespoulous, J. (2006) Working memory performance in expert and novice interpreters. *Interpreting*, 8 (1), pp. 1-23.
  128. Korpál, P. (2012) Omission in simultaneous interpreting as a deliberate act. *Translation Research Projects* 4, , pp. 103-111.
  129. Korpál, P. (2015) Eye-tracking in Translation and Interpreting Studies: The growing popularity and methodological problems. In: B. L and D. M, eds. *Assessing audiovisual translation*. Peterlang, pp. 199-212.
  130. Korpál, P. (2016) Interpreting as a stressful activity: Physiological measures of stress in simultaneous interpreting. *Poznan Studies in Contemporary Linguistics*, 52 (2), pp. 297- 316.
  131. Kraviarová, Z. (2013). Assessing Source Speech Difficulty for Research Purposes. In FORUM. *Revue Internationale d'interprétation et de traduction/International Journal of Interpretation and Translation*. 11(1), 51-75.
  132. Kucera, H. and Francis, W.N. (1967) *Computational Analysis of Present-Day American English*.
  133. Kumcu, A. (2011) *Visual focal loci in simultaneous interpreting*. Master, Hacettepe University Graduate School of Social Sciences.
  134. Lader, M. and Montagu, J., D. (1962) Skin Conductance During Stress and Relaxation. *The Journal of Nervous and Mental Disease*, 135 (4), pp. 259-270.
  135. Laeng, B., Sirois, S. and Gredebäck, G. (2012) Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7 (1), pp. 18-27.
  136. Lamberger-Felber, H. (2001). Text-oriented research into interpreting-Examples from a case-study. *HERMES-Journal of Language and Communication in Business*, 14(26), 39-64.

137. Lambert, S. (1988). Information processing among conference interpreters: A test of the depth-of-processing hypothesis. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 33(3), 377-387.
138. Lambert, S. (2004) Shared Attention during Sight Translation, Sight Interpretation and Simultaneous Interpretation. *Meta*, 49 (2), pp. 294-306.
139. Lederer, M. (1981). *La traduction simultanée: expérience et théorie* (Vol. 3). Lettres modernes.
140. Lee, T. (1999) Speech proportion and accuracy in English into Korean simultaneous interpretation. *Meta*, 44 (4), pp. 60-65.
141. Li, D. (2019) *The Effect of Syntactic Difficulty on Cognitive Effort in English–Chinese Sight Translation: An Eye Tracking Study*. Master, University of Stirling.
142. Lin, Y., Lv, Q. and Liang, J. (2018) Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Predicting Fluency with Language Proficiency, Working Memory, and Directionality in Simultaneous Interpreting*, 9, pp. 1-13.
143. Liu, M. (2001) *Expertise in simultaneous interpreting: a working memory analysis*. Doctor of Philosophy, The University of Texas at Austin.
144. Liu, M. and Chiu, Y. (2009) Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting*, 11 (2), pp. 244-266.
145. Liu, M., Schallert, D. and Carroll, P. (2004) Working memory and expertise in simultaneous interpreting. *International Journal of Research and Practice in Interpreting*, 6 (1), pp. 19-42.
146. Liu, M. (2013) Design and Analysis of Taiwan's Interpretation Certification Examination. In: *Assessment Issues in Language Translation and Interpreting*. Peter Lang, pp. 163.
147. Lochner, M., Duenser, A. and Sarker, S. (2019) Trust and Cognitive Load in semi-automated UAV operation. *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. Fremantle WA Australia, December 2 - 5. New York, United States: Association for Computing Machinery , pp. 437–441.
148. Loewenfeld, I., E. (1958) Mechanisms of reflex dilatation of the pupil. Historical review and experimental analysis. *Documenta Ophthalmologica*, 12, pp. 185-448.
149. Mackintosh, J. (1983) Relay Interpretation: An Expoloratory Study. MA dissertation, Birkbeck College, University of London.
150. Marks, C, B., Doctorow, M, J. and Wittrock, M, C. (1974) Word frequency and reading comprehension. *The Journal of Educational Research*, 67 (6), pp. 259–262.

151. Ma, X. (2021) Coping with syntactic complexity in English-Chinese sight translation by translation and interpreting students: An eye-tracking investigation. *Across Languages and Cultures*, 22 (2), pp. 192-213.
152. Ma, X. and Li, D. (2021) A cognitive investigation of ‘chunking’ and ‘reordering’ for coping with word- order asymmetry in English-to-Chinese sight translation: Evidence from an eye-tracking study. *International Journal of Research and Practice in Interpreting*, 23 (2), pp. 192-221.
153. Ma, X., Li, D. and Hsu, Y. (2021) Exploring the impact of word order asymmetry on cognitive load during Chinese-English sight translation Evidence from eye-movement data. *Target. International Journal of Translation Studies*, 33 (1), pp. 103-131.
154. Macnamara, N. and Conway, A. (2016) Working memory capacity as a predictor of simultaneous language interpreting performance. *Journal of Applied Research in Memory and Cognition*, 5 (4), pp. 434–444.
155. Macnamara, N., Conway, A. and Kegl, J. (2011) Domain-general cognitive abilities and simultaneous interpreting skill. *Interpreting*, 13 (1), pp. 121–142.
156. Matasov, R. (2017) Nuremberg: The Trial of Six Million Words. *Aiic*, .
157. Mathôt, S. (2018) Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, 1 (1), pp. 1-23.
158. Mayer, R.E. and Moreno, R. (2010) Techniques That Reduce Extraneous Cognitive Load and Manage Intrinsic Cognitive Load during Multimedia Learning. In: *Cognitive Load Theory*. Cambridge University PressApr 26, pp. 131-152.
159. Mazza, C. (2001) Numbers in Simultaneous Interpretation. *The Interpreters’ Newsletter*, 11, pp. 87–104.
160. McCleary, R., A. (1950) The nature of the galvanic skin response. *Psychological Bulletin*, 47 (2), pp. 97-117.
161. McKiever, B., L. and Mundy-Castle, A., C. (1953) The Psychophysiological Significance of The Galvanic Skin Response. *Journal of Experimental Psychology*, 46 (1), pp. 15-24.
162. Mead, P. (2015) Input Variables. in Pöchhacker, F. (ed). *Routledge encyclopedia of interpreting studies*. Routledge, 191-192.
163. Merwe, A., B. and Theron, P., A. (1947) A new method of measuring emotional stability. *The Journal of General Psychology*, 37, pp. 109-123.

164. Michael, E., Keller, T., Carpenter, P. and Marcel, J. (2001) fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, 13 (4), pp. 239–252.
165. Mikkelsen, H. (1993) *Introduction to sight translation from the interpreter's edge*. USA: ACEBO.
166. Miller, G., A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63 (2), pp. 81-97.
167. Mizuno, A. (2017) Simultaneous interpreting and cognitive constraints. *Bulletin of College of Literature*, 58, pp. 1-28.
168. Monod, H. and Kapitaniak, B. (1999) *Ergonomie* . Paris: Masson.
169. Montagu, J., D. and Coles, E., M. (1966) MECHANISM AND MEASUREMENT OF THE GALVANIC SKIN RESPONSE. *Psychological Bulletin*, 65 (5), pp. 261-279.
170. NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (1980) *NASA TLX LOAD INDEX*. NASA. Available: <https://humansystems.arc.nasa.gov/groups/TLX/>
171. Nobre, A., Rodrigues, J., Sbicigo, J., Piccolo, L., Zortea, M., Junior, S. and Salles, J. (2013) Tasks for assessment of the episodic buffer: A systematic review. *Psychology and Neuroscience*, 3 (10), pp. 331-343.
172. Nolan, J. (2005) *Interpretation: Techniques and exercises*. *Multilingual Matters*. Multilingual Matters.
173. Nourbakhsh, N., Wang, Y. and Chen, F. (2013) GSR and Blink Features for Cognitive Load Classification . In: P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson and M. Winckler, ed. *Human-Computer Interaction – INTERACT 2013*. South Africa, September 2-6, 2013. Springer , pp. 159-166.
174. Nourbakhsh, N., Wang, Y., Chen, F. and Calvo, R.A. (2012) Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In: V. Farrell, G. Farrell, C. Chua, W. Huang, R. Vasa and C. Woodward, ed. *Proceedings of the 24th Australian Computer-Human Interaction Conference*. Melbourne Australia, November 26 - 30. New York, United States: Association for Computing Machinery , pp. 420–423.
175. Oberauer, K. (2004) The measurement of working memory capacity. In: O. Wilhelm and R. Engle, eds. *Handbook of understanding and measuring intelligence*. Sage Publications, pp. 393–407.
176. Oléron, P., & Nanpon, H. (1965). Recherches sur la traduction simultanée [Studies of simultaneous translation]. *Journal de Psychologie Normale et Pathologique*, 62(1), 73-94.

177. Orru, G. and Longo, L. The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review. In: *Human Mental Workload: Models and Applications*. Cham: Springer International Publishing, pp. 23-48.
178. Paas, F., Tuovinen, J.E., Tabbers, H. and Van Gerven, P.W.M. (2003) Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38 (1), pp. 63-71.
179. Padilla, P., Bajo, M., Canas, J. and Padilla, F. (2005) Cognitive processes of memory in simultaneous interpretation. In: J. Tömmola, ed. *Topics in interpreting research*. Turku: Painosalama, pp. 61-72.
180. Pashler, H. (1994) Dual-Task Interference in Simple Tasks. *Psychological Bulletin*, 116 (2), pp. 220-244.
181. Patterson, K. and Shewell, C. (1987) Speak and spell: dissociations and word-class effects. In: M. Coltheart et al., ed. *The cognitive neuropsychology of language*. London: Erlbaum, pp. 273-294.
182. Pavlović, N., & Jensen, K. (2009). Eye tracking translation directionality. *Translation research projects*, 2, 93-109.
183. Pio, S. (2003) The Relation Between St Delivery Rate and Quality In Simultaneous Interpretation. *The Interpreters' Newsletter*, , pp. 69-100.
184. Plevóets, K. and Defrancq, B. (2016) The effect of informational load on disfluencies in interpreting. A corpus-based regression analysis. *Translation and Interpreting Studies*, 11 (2), pp. 202-224.
185. Plevóets, K. and Defrancq, B. (2018) The cognitive load of interpreters in the European Parliament A corpus-based study of predictors for the disfluency uh(m) Koen. *International Journal of Research and Practice in Interpreting*, 20 (1), pp. 1-32.
186. Pöchhacker, F. (2001) Quality Assessment in Conference and Community Interpreting. *Meta*, 46 (2), pp. 410-425.
187. Pöchhacker, F. (2007). Coping with culture in media interpreting. *Perspectives: Studies on Translatology*, 15(2), 123-142.
188. Pöchhacker, F. (2009) Conference interpreting: Surveying the profession. *Translation and Interpreting Studies*, 4 (2), pp. 172-186.
189. Pochhacker, F. (2016) *Introducing interpreting studies*. Second Edition ed. London ; New York: Routledge.

190. Pointe La, L.B. and Engle, R.W. (1990) Simple and Complex Word Spans as Measures of Working Memory Capacity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 16 (6), pp. 1118-1133.
191. Pym, A. (2009) On omission in simultaneous interpreting: Risk analysis of a hidden effort. In: H. Gerzymisch-Arbogast, ed. *Efforts and Models in Interpreting and Translation Research*. Amsterdam/Philadelphia: John Benjamins, pp. 83-105.
192. Rayner, K. and Duffy, S. (1986) Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14 (3), pp. 191–201.
193. Riccardi, A., Marinuzzi, G. and Zecchin, S. (1998) INTERPRETATION AND STRESS. *The Interpreter's Newsletter*, 8, pp. 93-106.
194. Riccardi, A. (2015) Speech Rate. in Pöchhacker, F. (ed). Routledge encyclopedia of interpreting
195. Richardson, D.C. and Spivey, M.J. (2004) Eye tracking: Characteristics and methods. In: W. G and B. G, eds. *Encyclopedia of biomaterials and biomedical engineering*. New York: Marcel Dekker, pp. 568–572.
196. Rosen, V. and Engle, R. (1997) Forward and backward serial recall. *Intelligence (Norwood)*, 24 (1), pp. 37.
197. Russo, M. (2011) Aptitude testing over the years. *Interpreting*, 13 (1), pp. 6–30.
198. Ryder, R.J. and Slater, W.H. (1988) The Relationship Between Word Frequency and Word Knowledge. *The Journal of Educational Research (Washington, D.C.)*, 81 (5), pp. 312-317.
199. Sawyer, B. (2004) *Fundamental Aspects of Interpreter Education: Curriculum and Assessment*. 1 ed. Amsterdam: John Benjamins.
200. Schultheis, H. and Jameson, A. (2004) Assessing Cognitive Load in Adaptive Hypermedia Systems: Physiological and Behavioral Methods. In: W. Nejdl and P. De Bra, ed. *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Berlin, pp. 225–234.
201. Seeber, K. (2011) Cognitive load in simultaneous interpreting: Existing theories— new models. *Interpreting*, 13 (2), pp. 176-204.
202. Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, 16(2), 228-242.
203. Seeber, K. (2013) Cognitive load in simultaneous interpreting: Measures and methods. *Target*, 25 (1), pp. 18\_32.

204. Seeber, K. G. (2015) SI with text. In Mikkelsen, H., & Jourdenais, R. (eds.). *The Routledge handbook of interpreting*. Routledge, 87-88.
205. Seleskovitch, D. (1978) Language and Cognition. *Language Interpretation and Communication*, 6, pp. 333-341.
206. Seleskovitch, D. (1978). Interpreting for international conferences: Problems of language and communication. Pen & Booth
207. Sergio, V. (1991) Tag: memory Shut up & listen. *Annual Meeting of the First Language International Conference* &nbsp; Denmark.
208. Setton, R. (1999) *Simultaneous interpretation: A cognitive-pragmatic analysis*. Amsterdam: John Benjamins.
209. Shallice, T. (1987) Impairments of semantic processing: multiple dissociations. In: M. Coltheart et al., ed. *The cognitive neuropsychology of language*. London: Erlbaum, pp. 111–127.
210. Sharmin, S., Oleg, S., Kari-Jouko, R. and Arnt, J. (2008) Where and for How Long Do Translators Look at the Screen while Translating? In: G. Susanne et al., ed. *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*. Frederiksberg: Samfundslitteratur, pp. 31–51.
211. Shao, Z. and Chai, M. (2021) The effect of cognitive load on simultaneous interpreting performance: an empirical study at the local level. *Perspectives: Studies in Translation Theory and Practice*, 29 (5), pp. 778-794.
212. Sharma, M., Kacker, S. and Sharma, M. (2016) A Brief Introduction and Review on Galvanic Skin Response. *International Journal of Medical Research Professionals*, 2 (6), pp. 13-17.
213. Shi, Y., Rioz, N., Taob, R., Choi, E. and Chen, F. (2007) Galvanic skin response (GSR) as an index of cognitive load. *CHI '07 Extended Abstracts on Human Factors in Computing Systems*. San Jose CA USA, 28 April - 3 May. New York, United States: Association for Computing Machinery, pp. 2651–2656.
214. Shlesinger, M., Le Féal, K.D., Kurz, I., Mack, G., Cattaruzza, L., Nilsson, A., Niska, H., Pöchhacker, F. and Viezzi, M. (1997) Quality in Simultaneous Interpreting. In: *Conference Interpreting: Current Trends in Research*. Benjamins, pp. 123.
215. Shlesinger, M. (2000). Interpreting as a Cognitive Process: How can we know what really happens? In Jääskeläinen, R., & Tirkkonen-Condit, S. (eds). *Tapping and mapping the processes of translation and interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing, 3-16.
216. Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreters' Newsletter*, 12, 37-49.

217. Shreve, G., Lacruz, I. and Angelone, E. (2010) Cognitive effort, syntactic disruption, and visual interference in a sight translation task. In: G. Shreve and E. Angelone, eds. *Translation and Cognition*. John Benjamins Publishing Company, pp. 63–84.
218. Sjørup, A. C. (2013). Cognitive effort in metaphor translation. An eye-tracking and key-logging study (doctoral dissertation), Copenhagen Business School.
219. Song, S. and Li, D. (2020) The Predicting Power of Cognitive Fluency for the Development of Utterance Fluency in Simultaneous Interpreting. *Frontiers in Psychology*, 11, pp. 1-13.
220. Song, Z. (2010) Skill transfer from sight translation to simultaneous interpreting: a case study of effective teaching technique. *International Journal of Interpreter Education*, 2, 20 July
221. Sun, S. (2015) Measuring translation difficulty: Theoretical and methodological considerations. *Across Languages and Cultures*, 16 (1), pp. 29-54.
222. Su, W. (2019) *Eye Tracking Translation Styles in Chinese-English Sight Translation*. Doctor of Philosophy in Linguistics, University of Macau.
223. Su, W. and Li, D. (2019) Identifying translation problems in English-Chinese sight translation: An eye-tracking experiment. *Translation and Interpreting Studies*, 14 (1), pp. 110-134.
224. Swanson, J.M., Kraemer, H.C., Hinshaw, S.P., Arnold, L.E., Conners, C.K., Abikoff, H.B., Clevenger, W., Davies, M., Elliott, G.R., Greenhill, L.L., Hechtman, L., Hoza, B., Jensen, P.S., March, J.S., Newcorn, J.H., Owens, E.B., Pelham, W.E., Schiller, E., Severe, J.B., Simpson, S., Vitiello, B., Wells, K., Wigal, T. And Wu, M. (2001) Clinical Relevance of the Primary Findings of the MTA: Success Rates Based on Severity of ADHD and ODD Symptoms at the End of Treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40 (2), pp. 168-179.
225. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.
226. Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3), 251-296.
227. Teixeira, T., Wedel, M. and Pieters, R. (2014) Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research*, 51 (2), pp. 144-159.
228. Theron, P., A. (1948) Peripheral vasomotor reactions as indices of basic emotional tension and lability. *Psychosom*, 10, pp. 335-346.
229. Tiselius, E. (2015) Accuracy. In: F. Pöchhacker, ed. *Routledge Encyclopedia of Interpreting Studies*. 1 ed. London: Routledge, pp. 3-4.
230. Tobii (2010) *Tobii Studio User Manual*.

231. Tommola, J. and Heleva, M. (1998) Language direction and source text complexity effects on trainee performance in simultaneous interpreting. In: L. Bowker, M. Cronin, D. Kenny, J. Pearson, ed. *Unity in diversity*, Manchester: St. Jerome, pp. 177-186.
232. Tommola, J. and Niemi, P. (1986) Mental load in simultaneous interpreting: An online pilot study. In: L.S. Evensen, ed. *Nordic research in text linguistics and discourse analysis: Papers from the first NORDTEXT Symposium, Trondheim, August 15-18, 1985*. Oslo: Tapir, pp. 171-184.
233. Tommola, J. and Hyona, J. (1990) Mental Load in Listening, Speech Shadowing and Simultaneous Interpreting: A Pupillometric Study. *University of Turku*, , pp. 1-11.
234. Tzou, Y., Eslami, Z. and Chen, H. (2012) 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), pp. 213-227.
235. Viaggio, S. (1995) The Praise of Sight Translation (And Squeezing the Last Drop Thereout of) .
236. Viaggio, S. (2002) The quest for optimal relevance: The need to equip students with a pragmatic compass. In: M. Garzone and M. Viezzi, eds. *Interpreting in the 21st century: challenges and opportunities*. Amsterdam and Philadelphia: John Benjamins, pp. 229-244.
237. Viezzi, M. (1989) Information Retention as a Parameter for the Comparison of Sight Translation and Simultaneous Interpretation: An Experimental Study. *The Interpreters' Newsletter*, (2), pp. 65-69.
238. Visson, L. (2005) Simultaneous interpretation: Language and cultural difference. In: S. Bermann and M. Wood, eds. *Nation, Language, and the Ethics of Translation*. Princeton: Princeton: Princeton University Press.
239. Wadensjö, C. (1998) *Interpreting as interaction*. 1. publ. ed. London [u.a.]: Longman.
240. Wang, C., Baird, T., Huang, J., Coutinho, D., J., Brien, C., D. and Munoz, P., D. (2018) Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task. *Frontiers in Neurology*, 9.
241. Wang, Y. (2017) Translating Linguistic Metaphors in Both Directions: A Process-Oriented Study on English-Chinese Translation. Doctor of Philosophy, Durham University.
242. Welford, T. (1952) The 'Psychological Refractory Period' and the Timing of High Speed Performance — A Review and a Theory. *British Journal of Psychology*, 43, pp. 2-19.
243. Weng, Y., Zheng, B. and Dong, Y. (2022) Time pressure in translation. *Target. International Journal of Translation Studies*, 34 (4), pp. 601-626.

244. Wickens, C.D. (2002) Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3 (2), pp. 159-177.
245. Williams, J. and Chesterman, A. (2002) *The Map: A Beginner's Guide to Doing Research in Translation Studies*. 1 ed. Routledge.
246. Yang, S. (2019) *Investigating the Effect of Speech Rate on the Cognitive Load in Simultaneous Interpreting with Text*. PhD, University of Macau.
247. Zhang, W. (2012) 工作记忆与口译技能在同声传译中的作用与影响. *Foreign Language and Teaching and Research*, 5, pp. 751-764.
248. Zwischenberger, C. (2010) Quality criteria in simultaneous interpreting: an international vs. a national view. *The Interpreters' Newsletter*, 15, pp. 127–142.

# Appendices

## Appendix 1 Online Study (Experiment One) Consent Form

Version date: January 2020 [ESSENTIAL: Ensure the information sheet is dated and has a version number]  
[Funder logo if appropriate]



### Consent Form

GUEP/NICR Approval Number [Insert]

**Research Project Title: Interpreters' Cognitive Load Research**

|                                                                                                                                                                                                                                                                                                                    | Please initial box |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| I confirm that I have read and understood the Participant Information Sheet explaining the above research project and I have had the opportunity to ask questions about the project                                                                                                                                |                    |
| I understand that I have a chance to £10 amazon voucher and that I am free to withdraw at any time during the study and withdraw my data within two weeks of testing without giving a reason, and without any penalty. I understand that beyond that date it may not be possible to remove my data from the study. |                    |
| I understand that my responses will be kept anonymous and I give permission for members of the research team to have access to my anonymised responses.                                                                                                                                                            |                    |
| I consent to being audio recorded.                                                                                                                                                                                                                                                                                 |                    |
| I understand how audio will be used in research outputs. I am aware that I will not be named in any research outputs but I could be identified by people I know through the stories I tell.                                                                                                                        |                    |
| I agree for research data collected in the study to be given to researchers, including those working outside the EU to be used in other research studies. I understand that any data that leave the research group will be fully anonymised so that I cannot be identified.                                        |                    |
| I agree to take part in this study                                                                                                                                                                                                                                                                                 |                    |

## Appendix 2 Lab Study (Experiment Two) Consent Form

Version date: January 2020 [ESSENTIAL: Ensure the information sheet is dated and has a version number]  
[Funder logo if appropriate]



### Electronic Consent Form

GUEP/NICR Approval Number [Insert]

**Research Project Title: Cognitive Differences Between Sight Translation and Simultaneous Interpreting**

|                                                                                                                                                                                                                                                                                                                | Please initial box |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| I confirm that I have read and understood the Participant Information Sheet explaining the above research project and I have had the opportunity to ask questions about the project                                                                                                                            |                    |
| I understand that I can receive £10 after this test and that I am free to withdraw at any time during the study and withdraw my data within two weeks of testing without giving a reason, and without any penalty. I understand that beyond that date it may not be possible to remove my data from the study. |                    |
| I understand that my responses will be kept anonymous and I give permission for members of the research team to have access to my anonymised responses.                                                                                                                                                        |                    |
| I consent to being audio recorded.                                                                                                                                                                                                                                                                             |                    |
| I understand how audio will be used in research outputs. I am aware that I will not be named in any research outputs but I could be identified by people I know through the stories I tell.                                                                                                                    |                    |
| I agree for research data collected in the study to be given to researchers, including those working outside the EU to be used in other research studies. I understand that any data that leave the research group will be fully anonymised so that I cannot be identified.                                    |                    |
| I agree to take part in this study                                                                                                                                                                                                                                                                             |                    |

## Appendix 3 Online Experiment Tutorial (in Chinese)

### 实验教程

非常感谢您参加这个实验，在实验开始前，请您花一点时间阅读一下本教程，这会让实验进行的更加顺畅！

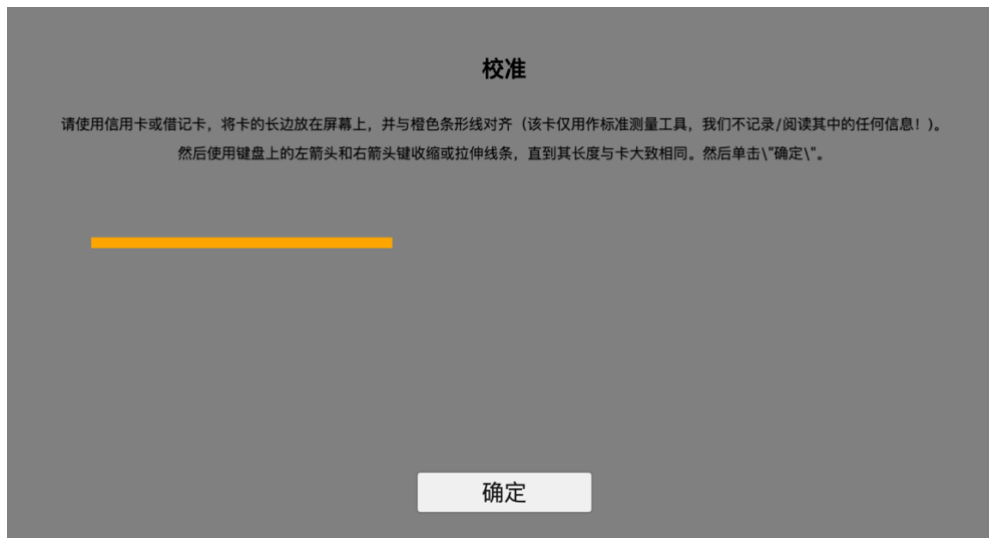
1. 进行本实验，您需要一台**电脑** (iPad 无法进行最后一项测试)、一个**鼠标**和一副**耳机**，请您一定**全程佩戴耳机**完成实验，在翻译的时候请您尽量贴近耳机的麦克风。请务必使用 **Chrome** 浏览器打开实验链接。谢谢。
2. 正式实验前，您会看到下面这个页面（如下图），如果您觉得设置为全屏很麻烦的话也没关系，不用根据下图的要求操作，按照您喜欢的窗口大小进行实验就可以了：），如果遇到有的页面看不到底部的 **START** 或是 **NEXT**，您只要往下拉一下页面就行了。

**开始之前：**

- 1) 将浏览窗口设置为全屏（您可以通过在Windows和Linux上按（F11）或在Mac上按组合键（CTRL + CMD + f）来执行此操作。如果以上方法不行，在Chrome浏览器上，则可以转到 查看菜单，然后点击全屏）
- 2) 尽量减少可能会影响您测试期间的表现的分心事物（比如电视，电话等）
- 3) 将自己从电脑显示器置于一臂长的距离空间，并尝试在整个实验过程中保持此距离

完成这些操作后，单击完成。

3. 实验开始前，还会出现让您“校准”的页面（如下图），您可以忽略页面上的要求，直接点击确定就可以了。



4. 实验开始后，您会看到让您打开麦克风的页面（如下图），因为需要对您的同传内容进行录音，所以请您点击“允许”。所有的实验结果都是匿名的，请您放心。



5. 接下来这个**很重要**，如果您看到校准声音的界面（如下图），请您**大声读出**页面上的句子，这是为了能精准的收入您好听的声音：）。

## 需要校准。

请大声读出这些句子：

There's nothing you can do that can't be done

Nothing you can sing that can't be sung

Nothing you can say but you can learn how to play the game

It's easy

校准将在几秒钟后结束。 3

6. 整个实验过程都会有详细的指引，请您按照步骤完成实验即可。
7. 视译和口译实验结束后，会显示数据正在上传，这时候实验还没有结束，请**不要关闭窗口**。
8. 接下来您会看到 end screen 的页面，点击“END”图标后，您会看到连接工作记忆（working memory）测试的页面（如下图）。请复制 **confirmation code**，也就是下图所示的 **A**；然后点击 **B**，也就是 **working memory test**。

Now, please access the final part of this test! That is working memory test!

The confirmation code for completing this experiment is:

aizxOOGJm07m2KPI **A**

Please click the "working memory test" button **working memory test** **B**

数据已保存。

9. 在点击 **working memory test** 后，您的电脑会打开一个新的页面（如下图），请一定**不要关闭**之前进行翻译测试的页面（您桌面上的窗口请一个都不要关闭，谢谢），然后请选择“点击此处开始调查”。

## 关于本次调查

感谢各位参加这个实验，这是本实验的最后一个环节——working memory test.

下面我们就开始吧！

## 联系信息

关于研究人员:

邮件地址: [duoduo.li@stir.ac.uk](mailto:duoduo.li@stir.ac.uk)

## 对你电脑的重要技术要求

你的浏览器估计是这个版本（括号里是版本号）：Chrome（91）  
你的浏览器支持本实验：)

本次实验需要用到一个键盘

[点击此处开始调查](#)

## Important data protection information

When you start, this survey will store your answers and browser information on the [PsyToolkit server](#). The responsibility for this survey rests entirely with the researcher(s) listed above. [Click here if you do not want to participate now.](#)

10. 请将复制好的 code 粘贴到下图的 **participant number** 即可，如果你刚刚忘记复制了也没事，回到刚才翻译测试的页面就可以找到 code 了。

Please type your participant number below

participant number:

请将上一个页面的 code 复制到这里

[点击按钮继续](#)

非常非常感谢看到这里的您，接下来，就请您开启实验吧，谢谢！！

## Appendix 4 NASA Task Load Index (TLX)

### ***NASA Task Load Index***

*Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.*

|      |      |      |
|------|------|------|
| Name | Task | Date |
|------|------|------|

Mental Demand      How mentally demanding was the task?

Very Low      Very High

Physical Demand      How physically demanding was the task?

Very Low      Very High

Temporal Demand      How hurried or rushed was the pace of the task?

Very Low      Very High

Performance      How successful were you in accomplishing what you were asked to do?

Perfect      Failure

Effort      How hard did you have to work to accomplish your level of performance?

Very Low      Very High

Frustration      How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low      Very High

## Appendix 5 Marking Criteria for Each Score Range

|                 |                                                                                                                                                                                                                      |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Marks 80% - 90% | Comprehensive and coherent conveyance of both the core arguments or narratives and associated details from the source text, with minimal to no inaccuracies or omissions.                                            |
| Marks 70% - 80% | Accurate conveyance of the core arguments or narratives from the source text; any inaccuracies are minor and do not distort the overall message.                                                                     |
| Marks 60% - 70% | Conveyance of the majority of the source text's core arguments or narratives with only minor inaccuracies, distortions, or omissions that do not significantly impact the holistic understanding of the text.        |
| Marks 50% - 60% | Successful conveyance of the main points of the source text, though there may be one significant inaccuracy, distortion, or omission or several minor ones that mildly impact the coherence of the conveyed message. |
| Marks 40% - 50% | Conveyance of the basic message of the source text, but significant inaccuracies or omissions substantially impact the accuracy of the conveyed message.                                                             |
| Marks below 40% | The interpretation lacks coherence, distorting the narrative significantly, with numerous inaccuracies or omissions.                                                                                                 |

## Appendix 6 Experiment One Results

### Accuracy Scores for Each Participant Under Each Condition

N: Normal Condition

D: Difficult Condition

F: Fast Condition

| Participant | ST (N) | ST (D) | ST (F) | SI (N) | SI (D) | SI (F) |
|-------------|--------|--------|--------|--------|--------|--------|
| 1           | 77.5   | 46.5   | 51.5   | 61.5   | 39     | 46     |
| 2           | 85     | 47     | 72     | 42     | 49     | 30.5   |
| 3           | 60.5   | 35     | 60.5   | 45     | 47     | 35     |
| 4           | 84     | 49     | 82.5   | 74     | 40     | 42     |
| 5           | 77     | 42.5   | 47     | 78     | 44     | 31.5   |
| 6           | 79     | 61.5   | 54     | 64     | 36.5   | 42.5   |
| 7           | 85     | 50     | 49     | 74.5   | 33     | 36     |
| 8           | 77     | 41     | 45     | 55.5   | 35     | 41     |
| 9           | 82.5   | 53.5   | 73     | 47     | 44.5   | 30     |
| 10          | 40     | 56.5   | 66     | 48     | 44.5   | 33     |
| 11          | 73     | 38     | 58     | 75.5   | 46     | 48.5   |
| 12          | 64.5   | 45     | 55     | 62     | 57     | 33     |
| 13          | 82.5   | 49     | 68.5   | 50.5   | 38     | 30     |
| 14          | 81     | 54     | 80     | 82     | 39.5   | 30     |
| 15          | 82     | 61     | 43     | 72     | 43.5   | 52     |
| 16          | 80     | 66     | 67.5   | 79     | 45     | 32     |
| 17          | 78.5   | 35     | 36.5   | 47.5   | 44     | 33.5   |
| 18          | 72     | 42     | 60.5   | 55     | 54     | 38     |
| 19          | 76     | 36     | 47     | 56     | 43.5   | 30.5   |
| 20          | 80     | 50     | 56.5   | 74.5   | 74.5   | 38     |
| 21          | 71     | 35     | 39.5   | 61     | 37     | 30     |
| 22          | 75.5   | 42     | 48     | 44.5   | 35     | 30     |
| 23          | 81     | 34     | 34     | 55     | 36     | 38     |
| 24          | 84     | 78     | 83     | 66     | 48.5   | 33     |

|    |      |      |      |      |    |      |
|----|------|------|------|------|----|------|
| 25 | 83.5 | 59   | 87   | 90   | 61 | 64   |
| 26 | 88   | 61.5 | 79.5 | 84   | 70 | 32.5 |
| 27 | 83   | 83   | 80   | 74.5 | 73 | 38.5 |
| 28 | 83   | 64.5 | 72.5 | 82.5 | 79 | 51   |
| 29 | 85   | 55   | 65.5 | 61   | 56 | 35   |
| 30 | 69   | 51   | 50   | 71.5 | 43 | 46   |
| 31 | 83   | 72.5 | 56   | 82   | 65 | 42   |
| 32 | 83.5 | 64   | 64   | 53.5 | 35 | 38.5 |

### Self-Assessment of Stress Level Score in Online Experiment

| Participant | ST (N) | ST (D) | ST (F) | SI (N) | SI (D) | SI (F) |
|-------------|--------|--------|--------|--------|--------|--------|
| 1           | 7      | 8      | 8      | 5      | 6      | 8      |
| 2           | 4      | 5      | 4      | 7      | 6      | 7      |
| 3           | 8      | 8      | 7      | 5      | 8      | 8      |
| 4           | 2      | 3      | 4      | 4      | 4      | 2      |
| 5           | 7      | 8      | 9      | 6      | 8      | 9      |
| 6           | 7      | 9      | 10     | 8      | 10     | 10     |
| 7           | 2      | 5      | 8      | 3      | 6      | 10     |
| 8           | 7      | 1      | 1      | 2      | 3      | 1      |
| 9           | 7      | 8      | 9      | 6      | 7      | 10     |
| 10          | 2      | 4      | 6      | 1      | 2      | 4      |
| 11          | 1      | 2      | 2      | 1      | 2      | 2      |
| 12          | 2      | 3      | 5      | 2      | 2      | 3      |
| 13          | 5      | 5      | 5      | 7      | 8      | 8      |
| 14          | 3      | 4      | 4      | 4      | 4      | 7      |
| 15          | 5      | 8      | 7      | 5      | 7      | 9      |
| 16          | 6      | 8      | 7      | 6      | 7      | 8      |
| 17          | 1      | 5      | 3      | 6      | 7      | 10     |
| 18          | 5      | 7      | 6      | 6      | 7      | 8      |
| 19          | 4      | 8      | 9      | 7      | 8      | 10     |
| 20          | 3      | 4      | 3      | 2      | 3      | 4      |
| 21          | 4      | 6      | 5      | 6      | 8      | 8      |
| 22          | 5      | 6      | 7      | 8      | 8      | 7      |

|    |   |   |   |   |   |    |
|----|---|---|---|---|---|----|
| 23 | 3 | 7 | 6 | 6 | 9 | 10 |
| 24 | 1 | 3 | 4 | 5 | 7 | 8  |
| 25 | 2 | 4 | 3 | 6 | 6 | 7  |
| 26 | 1 | 6 | 6 | 6 | 6 | 9  |
| 27 | 6 | 5 | 5 | 3 | 6 | 8  |
| 28 | 2 | 3 | 3 | 2 | 3 | 5  |
| 29 | 5 | 6 | 7 | 7 | 7 | 8  |
| 30 | 6 | 6 | 6 | 6 | 6 | 6  |
| 31 | 7 | 8 | 8 | 8 | 8 | 8  |
| 32 | 5 | 8 | 8 | 7 | 7 | 9  |

## Appendix 7 Experiment Two Results

### Participants' Accuracy Score

| pp | Perf SI<br>NN | Perf SI<br>NF | Perf SI<br>DN | Perf SI<br>DF | Perf ST<br>NN | Perf ST<br>NF | Perf ST<br>DN | Perf ST<br>DF |
|----|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1  | 67            | 47            | 58.5          | 32            | 78            | 75.5          | 73            | 71            |
| 2  | 83.5          | 60.5          | 75.5          | 48            | 85.5          | 86            | 84.5          | 81            |
| 3  | 59.5          | 37.5          | 50            | 30            | 74            | 71.5          | 67            | 62.5          |
| 4  | 73.5          | 48.5          | 68            | 43.5          | 80.5          | 79.5          | 80            | 72.5          |
| 5  | 76            | 53.5          | 69            | 46            | 83.5          | 81.5          | 81.5          | 73            |
| 6  | 73            | 48.5          | 64.5          | 38            | 79.5          | 78.5          | 79.5          | 72            |
| 7  | 58.5          | 37            | 42.5          | 30            | 74            | 69.5          | 65            | 62            |
| 8  | 69.5          | 48            | 60.5          | 38            | 79            | 78            | 73.5          | 71.5          |
| 9  | 87.5          | 67.5          | 77.5          | 51.5          | 86.5          | 87.5          | 87.5          | 82            |
| 10 | 96            | 92            | 92            | 57.5          | 93            | 90.5          | 90.5          | 94.5          |
| 11 | 83.5          | 66            | 76.5          | 48.5          | 86.5          | 87            | 86            | 81            |
| 12 | 77            | 54.5          | 69.5          | 47            | 85            | 82.5          | 82            | 76            |
| 13 | 89            | 72.5          | 79            | 54.5          | 88.5          | 88.5          | 89            | 82.5          |
| 14 | 92            | 74            | 80.5          | 54.5          | 91            | 90            | 90            | 94            |
| 15 | 68            | 47.5          | 59.5          | 37            | 78            | 77.5          | 73.5          | 71            |
| 16 | 37            | 30            | 32            | 30            | 73            | 62            | 63            | 59            |
| 17 | 65.5          | 41.5          | 54.5          | 30            | 75            | 74.5          | 68.5          | 66.5          |
| 18 | 58.5          | 31.5          | 37            | 30            | 73            | 65.5          | 63.5          | 62            |
| 19 | 63.5          | 40.5          | 53.5          | 30            | 74.5          | 74            | 68            | 63            |
| 20 | 66.5          | 41.5          | 56            | 31            | 77.5          | 74.5          | 69.5          | 70.5          |
| 21 | 75            | 51.5          | 68.5          | 46            | 82.5          | 81.5          | 80.5          | 72.5          |
| 22 | 83            | 57.5          | 71.5          | 47.5          | 85            | 83.5          | 84            | 81            |

### Participants' Pupil Size Value

| <b>pp</b> | <b>Pupil<br/>Size SI<br/>NN</b> | <b>Pupil<br/>Size SI<br/>NF</b> | <b>Pupil<br/>Size SI<br/>DN</b> | <b>Pupil<br/>Size SI<br/>DF</b> | <b>Pupil<br/>Size ST<br/>NN</b> | <b>Pupil<br/>Size ST<br/>NF</b> | <b>Pupil<br/>Size ST<br/>DN</b> | <b>Pupil<br/>Size ST<br/>DF</b> |
|-----------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| 1         | 5.778                           | 5.991                           | 5.853                           | 6.093                           | 5.014                           | 4.868                           | 5.701                           | 4.952                           |
| 2         | 4.662                           | 5.215                           | 4.915                           | 5.203                           | 4.406                           | 4.532                           | 5.159                           | 4.441                           |
| 3         | 5.932                           | 6.027                           | 5.834                           | 6.083                           | 4.802                           | 5.214                           | 5.840                           | 5.185                           |
| 4         | 6.172                           | 6.442                           | 6.230                           | 6.459                           | 5.290                           | 5.434                           | 4.757                           | 5.451                           |
| 5         | 4.232                           | 4.621                           | 4.585                           | 5.651                           | 5.601                           | 5.745                           | 3.662                           | 5.652                           |
| 6         | 5.778                           | 5.996                           | 5.853                           | 6.093                           | 4.868                           | 5.014                           | 4.882                           | 4.952                           |
| 7         | 5.328                           | 5.603                           | 5.077                           | 5.489                           | 4.510                           | 4.669                           | 5.359                           | 4.607                           |
| 8         | 4.924                           | 5.295                           | 5.395                           | 5.598                           | 4.856                           | 4.964                           | 5.584                           | 4.945                           |
| 9         | 4.386                           | 5.425                           | 4.986                           | 5.283                           | 4.655                           | 4.776                           | 4.872                           | 4.775                           |
| 10        | 4.355                           | 4.674                           | 4.130                           | 4.571                           | 3.678                           | 3.711                           | 3.680                           | 3.788                           |
| 11        | 4.833                           | 5.214                           | 5.791                           | 6.323                           | 5.261                           | 5.352                           | 4.857                           | 5.379                           |
| 12        | 5.333                           | 5.668                           | 5.319                           | 5.523                           | 4.654                           | 4.764                           | 3.864                           | 4.653                           |
| 13        | 5.292                           | 5.996                           | 5.511                           | 5.811                           | 4.847                           | 5.031                           | 3.916                           | 4.919                           |
| 14        | 5.238                           | 5.443                           | 5.293                           | 5.441                           | 4.436                           | 4.582                           | 3.843                           | 4.392                           |
| 15        | 3.865                           | 4.135                           | 4.222                           | 4.921                           | 3.850                           | 3.957                           | 5.327                           | 3.936                           |
| 16        | 6.807                           | 6.973                           | 6.674                           | 7.007                           | 5.670                           | 5.851                           | 4.426                           | 5.597                           |
| 17        | 5.303                           | 5.564                           | 5.241                           | 6.235                           | 5.565                           | 5.685                           | 5.307                           | 5.821                           |
| 18        | 4.572                           | 4.936                           | 4.489                           | 4.713                           | 3.770                           | 3.851                           | 4.660                           | 3.897                           |
| 19        | 4.054                           | 4.156                           | 4.089                           | 4.355                           | 3.873                           | 3.904                           | 3.806                           | 3.932                           |
| 20        | 4.690                           | 4.732                           | 3.737                           | 4.623                           | 3.679                           | 3.797                           | 4.853                           | 3.790                           |
| 21        | 6.300                           | 6.222                           | 5.774                           | 6.253                           | 5.431                           | 5.456                           | 4.568                           | 5.487                           |
| 22        | 4.475                           | 4.636                           | 4.531                           | 4.809                           | 3.947                           | 3.961                           | 4.489                           | 3.939                           |

### Participates' GSR Value

| pp | GSR<br>SI NN | GSR<br>SI NF | GSR<br>SI DN | GSR<br>SI DF | GSR<br>ST NN | GSR<br>ST NF | GSR<br>ST DN | GSR<br>ST DF |
|----|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1  | 0.079        | 0.164        | 0.190        | 0.200        | 0.218        | 0.256        | 0.275        | 0.326        |
| 2  | 0.522        | 0.625        | 0.506        | 0.466        | 0.497        | 0.605        | 0.702        | 0.561        |
| 3  | 0.368        | 0.399        | 0.317        | 0.371        | 0.421        | 0.454        | 0.470        | 0.431        |
| 4  | 0.081        | 0.111        | 0.122        | 0.189        | 0.278        | 0.323        | 0.270        | 0.282        |
| 5  | 0.499        | 0.517        | 0.548        | 0.481        | 0.438        | 0.422        | 0.425        | 0.435        |
| 6  | 0.873        | 0.934        | 0.824        | 1.218        | 0.781        | 0.778        | 0.776        | 0.811        |
| 7  | 0.401        | 0.296        | 0.374        | 0.298        | 0.402        | 0.445        | 0.015        | 0.015        |
| 8  | 0.826        | 0.802        | 0.961        | 0.886        | 0.978        | 1.123        | 0.936        | 0.923        |
| 9  | 1.128        | 0.988        | 1.024        | 1.206        | 1.400        | 1.497        | 1.484        | 1.318        |
| 10 | 0.803        | 0.953        | 0.827        | 0.815        | 1.030        | 1.023        | 0.955        | 0.947        |
| 11 | 0.856        | 0.817        | 0.810        | 0.945        | 0.950        | 0.963        | 1.013        | 0.993        |
| 12 | 2.271        | 2.347        | 2.455        | 2.213        | 2.469        | 2.422        | 2.564        | 2.462        |
| 13 | 1.487        | 1.702        | 1.440        | 1.288        | 1.776        | 1.825        | 1.788        | 1.741        |
| 14 | 0.480        | 0.319        | 0.369        | 0.434        | 0.555        | 0.519        | 0.519        | 0.585        |
| 15 | 1.422        | 1.735        | 1.397        | 1.503        | 1.039        | 0.908        | 1.187        | 1.404        |
| 16 | 1.721        | 1.760        | 1.705        | 1.756        | 1.720        | 1.658        | 1.659        | 1.535        |
| 17 | 1.523        | 1.839        | 1.559        | 1.555        | 1.318        | 1.019        | 1.121        | 1.384        |
| 18 | 2.638        | 2.638        | 2.617        | 2.769        | 2.258        | 2.045        | 2.405        | 2.518        |
| 19 | 0.565        | 0.543        | 0.657        | 0.545        | 0.349        | 0.183        | 0.162        | 0.212        |
| 20 | 0.688        | 0.947        | 0.692        | 0.747        | 0.528        | 0.425        | 0.537        | 0.482        |
| 21 | 0.610        | 0.882        | 0.550        | 0.553        | 0.112        | 0.249        | 0.360        | 0.191        |
| 22 | 0.013        | 0.012        | 0.011        | 0.015        | 0.016        | 0.017        | 0.024        | 0.036        |