An investigation of subtitles as learning support in university education

By

Wing Shan Chan

MRes, Macquarie University, 2016

MT&I, Macquarie University, 2015

GDipMusThy, University of Technology, Sydney, 1999

BSc, University of New South Wales, 1996

Faculty of Medicine, Health and Human Sciences

Department of Linguistics

Macquarie University

This thesis is presented for the fulfilment of the degree of Doctor of Philosophy Linguistics

Sydney, NSW, Australia

9 July 2020

TABLE OF CONTENT

Abstract

Past literature on the benefits of subtitles has yielded contradictory results as variables such as language proficiency, language history, subtitle language, and subtitle characteristics have not been fully and consistently controlled. Furthermore, there is little research on the actual processing and impact of subtitles as learning support for first and second language learners in academic contexts – an evidently growing area of need. Using a mixed-methods approach, the current study aims to investigate the processing and resultant impact of subtitles on English- and Chinese-speaking students in terms of their comprehension and cognitive load. Experiment 1 (n=103) of the study was a five-week data collection in an authentic university classroom and revealed that English-speaking participants had lower self-reported cognitive load in reading foreign language subtitles while the presence of subtitles had no impact on Chinese-speaking participants. Experiment 2 (n=70) was a laboratory-based experiment replicating Experiment 1 with a group of Chinese first-language students in a more controlled environment in which three conditions were examined: first-language subtitles, second-language subtitles, and no subtitles. First-language subtitles were found to result in improved comprehension. The current study adds further insight and complement existing research on the effectiveness of subtitles and subtitle language that second-language learners reading first-language subtitles improve performance in an academic context. The results provide a valuable implication on education pedagogy in terms of assisting second-language learners in achieving their highest potential academically without being disadvantaged by possible language barriers.

Statement of Candidate

I certify that the work in this thesis entitled "**An investigation of subtitles as learning support in university education**" has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any university or institution other than Macquarie University. I also certify that the thesis is an original piece of research and that it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis has gained ethics approval from Macquarie University (5201700903).

Signed:

Wing Shan Chan

9 July 2020

Acknowledgements

Foremost, I would like to express my deepest gratitude to my amazing supervisor Professor Jan-Louis Kruger, you have been a constant inspiration and support throughout my research journey, and your constant challenge to reach my full potential has pushed me to strive for new heights. I would also like to thank my adjunct supervisor Associate Professor Stephen Doherty, for being so calm and reassuring when I got stuck in my research process from time to time. Thank you also to Dr Pamela Humphreys, who approved my data collection at MUIC; Dr Prashan Karunaratne, who arranged the collaboration so I have the materials and data to work with in my project; and all the lecturers who teach ECON111 in MUIC, my data collection could not have been completed without your support and understanding. Thank you to Dr Cassi Liardet for being my mentor in the early stage of my candidature. Thank you to Dr Peter Humburg, who has become a crucial figure in helping me with the statistics. Thank you to everyone from the eye tracking lab who have helped me and guided me through the technical side of the technology. Thank you to Kittye, Alice and all students in Translation and Interpreting who took part in this research, your help has made my data collection much easier. Thank you to my PhD cohort in C3B, SiXin, Xiaomin, and Anh. Especially SiXin, thank you for your kindness and generosity, you are my best listener and the best babysitter ever! Thank you to Albert, who offered to read my draft; and Vince, who shared a few tips for final editing.

Thank you to my mum. Thank you for your effort in seeing who I really am, your constant support in my education has made this path possible!

To you, Neil, I cannot tell you how much I felt loved throughout this journey. My dream could not come true without your total support in every part of our life. Thank you for looking after our children when I needed to focus; thank you for believing in me when I was in doubt; thank you for being so gentle, sensitive and crazy; thank you for making me laugh! To all my little young ladies: Mary, Katie, Emma and Sophie, thank you for walking this road with me, and allowing me to just be me. Thank you for giving me so much laughter and tears.

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background and motivation

Although there has been a growing interest in the use of subtitling in education over the past four decades, this research has often focused mainly on the use of subtitles in language learning and for people with hearing loss. As online learning has become increasingly popular, the use of subtitling in educational video has also become a growing area of research interest to increase accessibility for a variety of population including people with hearing loss, foreign language speakers, and also first-language speakers. The use of educational subtitling has become essential as a result of increasing globalisation, it allows learners from around the world to learn through an online medium in a language of their choice; many students even travel and study overseas through a language that is not their first language. Generally, these students would have achieved minimum standard of language proficiency in order to study abroad, however, they still face a language barrier that could prevent these students from achieving their full potential. This educational phenomenon has created a growing research interest in the most effective ways to support these learners. Existing research has proven the benefits of subtitles, mostly in the context of language acquisition. In terms of reading and comprehending a second-language, methodologies used in past research mostly involved measuring reading and understanding text in static contexts, but rarely in dynamic contexts like subtitled video. There are a limited number of empirical studies that investigate the processing and impact of subtitles and subtitle language as learning support for first- (L1) and second- (L2) language learners in academic contexts, and therefore it is the goal of this study to explore this research area. This study adds to the understanding of the discipline and will potentially have important pedagogical implications in supporting academic success for second-language learners.

1.2 Current problems in inconclusive results on subtitle research

1.2.1 The benefits of subtitles

It has been showed that results from previous studies on the benefits of subtitles are inconclusive. Detail review of past research on the benefits of subtitles is presented in *Chapter 2* section 2.4. The wide variety of research contexts and designs with different focuses generate a wide spectrum of research outcomes, however, the fact that most audiovisual studies have imported methodologies from other well-established disciplines, such as psychology, psycholinguistics, and cognitive science rather haphazardly or selectively, makes it difficult to replicate studies as a result of inconsistent practices that lack standardisation in experimental protocols and frameworks (Orero et al., 2018).

1.2.2 Impact of subtitle language

In terms of research in subtitle language, various studies show that reading second-language subtitles is beneficial for second-language learners in the context of language acquisition (for example, Baltova, 1999; Bird & William, 2002; Danan, 2004; Hayati & Mohmedi, 2009; Mitterer & McQueen, 2009; Vanderplank, 1988); while Markham, Peter, and McCarthy (2001) showed that first-language subtitles is more beneficial resulting in better performance. On the other hand, the studies by Bisson, van Heuven, Conklin, and Tunney (2014) and Kruger, Hefer, and Matthew (2014) proved that subtitle language has no significant impact on performance. Detail review can be found in *Chapter 2*, section 2.4.2. However, these studies showed little information on the impact of second-language subtitles in content learning and their results are inconclusive. Furthermore, there is very limited research on the impact of subtitle language for second-language learners in an educational context.

1.2.3 Impact of redundancy effect

In educational psychology, a redundancy effect (Kalyuga & Sweller, 2014) is predicted when information is presented in more than one medium, which could potentially hold true for subtitled video. As subtitles are considered as redundant information in a multimedia learning environment that consists of visual image and audio information that present identical content to the subtitles, it stands to reason that the presence of subtitles would impose additional cognitive load. However, the extent to which this redundancy impacts the processing of subtitles is still unclear as previous studies also yielded inconsistent results. The review of this inconsistency is discussed in *Chapter 2*, section 2.4.3, and an alternative view proposed by Van der Zee, Admiraal, Paas, Saab, and Giesbers (2017) is presented in section 2.4.4.

1.2.4 Eye tracking methods

Eye tracking has become a popular measure in reading research investigating eye movement patterns and cognitive processing during reading, also in reading dynamic text in a multimedia learning environment. The advantage of eye tracking methods is its objective, direct and real-time measurements of cognitive processing (Kruger, 2016). However, due to the large amount of variation in content including the distribution of attention (e.g., Baltova, 1999; d'Ydewalle & de Bruycher, 2007; d'Ydewalle, Praet, Verfaillie, & Rensbergen., 1991; d'Ydewalle, van Rensbergen, & Pollet, 1987; Liao et al., 2020), subtitle language (e.g., Kruger, Hefer, & Matthew, 2013; Kruger, Hefer, & Matthew, 2014), one and two-line subtitles (e.g., d'Ydewalle & de Bruycher, 2007; d'Ydewalle et al., 1991; Kruger & Steyn, 2013), effort in subtitle reading (e.g., Kruger et al., 2014; Kruger & Steyn, 2013); language proficiency (e.g., Bird & Williams, 2002; Danan, 1992; Garza, 1991; Markham, 1999; van der Zee et al., 2017; Vanderplank, 1988) and other factors involved in learning such as individual learning preference (e.g., Homer, Plass & Black, 2008; Lincoln & Rademacher, 2006) and learners' motivation (Debue & van de Leemput, 2014), this is still an area in which there is a shortage of replicable large-scale studies that will allow for more robust conclusions on the impact of subtitles on learning. In particular, not much research has been conducted on the impact of first- and second-language subtitles on learning in an educational context involving eye movement patterns. This motivates the focus of the current study to investigate the impact of subtitles on students studying through the medium of a second-language with subtitles in either their first- or second-language, making use of eye tracking methods.

1.3 Thesis aim and frameworks

The current study aims to investigate the impact of subtitles and subtitle language on Chinese- and English-speaking learners in terms of performance and effort. Performance is measured as comprehension, and effort is measured through cognitive load (CL). Three research questions are formulated in order to achieve this goal via two experiments. Experiment 1, involving Chinese- and English-speaking learners, was conducted in a real-life classroom situation to answer Research Question 1: What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 or English L1 participants? Experiment 2, involving only Chinese-speaking learners, was conducted in a laboratory environment to answer Research Question 2: What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1? And Research Question 3: How do the presence of English and Chinese subtitles impact on the processing of audiovisual text by Chinese L1 students? These two experiments have been set up in different environments in order to answer these questions comprehensively. The design of the study used a mixed-methods approach consisting of the qualitative and quantitative analysis of eye tracking, performance and behavioural data. The hypotheses of this study were formulated based on a few conceptual frameworks including cognitive load theory by Sweller (2010, 2011), dual coding effect by Paivio (1990, 1991), redundancy effect by Kalyuga and Sweller (2014), lexical interference concept by Guichon and McLornan (2008), and automatic reading behaviour by d'Ydewalle and De Bruycker (2007). The detail of the hypotheses is outlined at the end of *Chapter 2*.

Chapter 2 Literature Review

2.1 Introduction

Subtitling has been widely used and researched in the context of movies and language learning, but there is limited empirical research on the impact of subtitles as learning support in content learning for second-language learners in an academic context. Based on the literature, limited information is available, and findings are inconclusive on the role of the language used in subtitles in the learning process. Furthermore, previous eye tracking studies used eye tracking mostly in laboratory context, limiting the ecological validity of findings in investigating the impact of subtitles and subtitle language on learning. The current study includes both a real-life (ecologically valid) situation and a controlled (laboratory-based) environment to investigate participants' subtitle reading behaviour. The goal of this study is to investigate whether subtitles support content learning and if so, whether first- or second-language subtitles has a bigger impact in supporting content learning in an educational context.

2.2 Cognitive Load Theory

Cognitive load theory (CTL) is an instructional theory based on an evolutionary view of the human cognitive architecture (Sweller, 2003, 2011; Sweller & Sweller, 2006). The evolutionary perspective on cognitive load theory draws on Geary's classification of knowledge into biologically primary and secondary knowledge (Geary, 2007, 2008). As explained by Sweller (2011), this classification is based on an assumption that humans have evolved to acquire certain types of information that can be learned effortlessly and unconsciously without extra instruction, such as speaking and listening. This knowledge is known as biologically primary knowledge. Another type of information is known as biologically secondary knowledge, which has not been required until relatively recently in the development of the species, therefore human beings have not yet evolved with any particular disposition to acquire that information. Biologically secondary knowledge such as reading and writing, as described by Sweller (2011), is important culturally and socially and is necessary to be learned in order to perform acceptably within any social context. Furthermore, instructional procedures are key in acquiring biologically secondary knowledge. Sweller (2003) suggested that the information structures that people have to engage with cognitively would impact the way humans evolve in their cognitive configuration. He further identified gaps in previous research on CLT which focused mainly on the organisation of the human cognitive architecture with very little attention to the way information was presented that have driven the evolution of that architecture. This explains the advantage of Geary's categorization of knowledge from an evolutionary perspective in the context of CLT as it is directly related to instructional procedures (Geary, 2007, 2008; Sweller, 2011).

According to Sweller (2004), the human cognitive architecture has a noticeable resemblance to the process of evolution by natural selection in terms of information storage. In any environment, effective variations will persist while ineffective variations will discontinue. Similarly, the human cognitive architecture consists of unlimited long-term memory that coordinates human cognitive activities, and a very limited working memory that tests the effectiveness of small variations to long-term memory. Any effective variations will alter and be stored in the long-term memory resulting in learning, while ineffective variations are lost and thus learning will not occur. Sweller (2003) indicated that the advantage of this evolutionary perspective is that the working memory utilises long-term memory as a backup to support effective processing. Working memory becomes optimal only when handling knowledge that has been learned previously and stored in long term memory. This process is also called 'schema automation' (Paas & Sweller, 2012, p. 29) and it is very limited when processing new information because there are no existing schemas, or 'central executive' (Sweller, 2003, p. 215), to co-ordinate novel information.

According to CLT, information can be assigned to a single unit within a cognitive schema that can be automated through learning by practicing or rehearsing, enabling the information to be transferred from novel stimuli into familiar stimuli which can be stored in long term memory. This allows stimuli to bypass working memory while information is mentally processed to free up space and thus increase working memory capacity (Paas, Tuovinen, Tabbers, & van Gerven, 2003). Since the working memory is so limited, only certain units can be processed at the same time. If too much information is presented at one time, cognitive overload could occur resulting in a decrease in learning. Cognitive overload could be measured through comprehension, recall, memory retention or recognition, and the reduction of scores is an indication of the presence of cognitive overload. One of the questions in subtitle reading concerns the vast amount of information presented simultaneously in different channels which could potentially result in overloading the cognitive processing and inevitably impact performance.

2.2.1 Three types of cognitive load

Cognitive load (CL) can be defined as a multidimensional construct that symbolises the load introduced to the learner's cognitive system when performing a particular task (Paas et al., 2003; Paas & van Merrienboer, 1994). Sweller (1994, 2003, 2010, 2011) proposed that element interactivity is the basic defining mechanism central to CLT and the cognitive load effects. According to Sweller (1994, p.304), 'an element is defined as any material that needs to be learned'. Sweller (2006, p. 13) also stated that 'when elements interact, they cannot be understood in isolation whereas non-interacting elements can be understood and learned independently of each other'. In other words, element interactivity is low when few elements interact; element interactivity is high when many elements interact (Sweller, 1994, 2006; Sweller & Chandler, 1994). Sweller (1994, p.304) further explained that 'the level of element interactivity …. refers to the extent to which the elements of a task can be meaningfully learned without having to learn the relations between any other elements'.

CLT differentiates 3 types of cognitive load: intrinsic load, extraneous load, and germane load (de Jong, 2010; Debue & van de Leemput, 2014; Leppink, Paas, van der Vleuten, van Gog, & van Merrienboer, 2013; Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014; Paas & Sweller, 2014; Paas et al., 2003; Sweller, 1988, 2010; Sweller, van Merrienboer, & Paas, 1998). Sweller (2010, 2011) defined these three types of cognitive load through the proposed concept of element interactivity. Intrinsic load is an interaction between the complexity of the material that needs to be learned and the learner's level of proficiency (Paas et al., 2003). Sweller (2010) proposed that the intrinsic cognitive load of a task relates to the level of element interactivity. He further suggested that the intrinsic complexity of the information to be processed by learners is fixed and unalterable unless altering what is learned or the learners' level of expertise (Sweller, 2010, 2011).

Extraneous load relates to the instructional format and the presentation of information that could potentially increase learner's overall cognitive load without improving learning, and badly designed instructional material imposes extraneous load (Debue & van de Leemput, 2014). Sweller (2011) indicated that extraneous load is also determined by element interactivity. He explained that by changing instructional procedures, the interacting elements that result in extraneous cognitive load can be reduced or eliminated.

Germane load is related to the process between constructs and automated schemas in long-term memory, and it is the remaining capacity that enables effective learning. According to Sweller (2010), germane load refers to the available resources in working memory that is necessary to deal with the element interactivity associated with intrinsic load, and learners have no control over germane load. Sweller (2010) argued that the three types of load are complementary, and the total cognitive load is the sum of all three types of load. If intrinsic or extraneous load is reduced, then germane load could potentially be increased without overloading working memory (see also, Ayres, 2006a, 2006b). In other words, an increase in germane load will result in learning enhancement as less working memory resources are being devoted to handling intrinsic and/or extraneous load. However, the results from the study of Debue and van de Leemput (2014) showed that germane load would not be directly associated with performance but would represent other cognitive processes that could improve learning, and that germane load could relate to learner motivation instead of task characteristics. If learners are motivated, they are more likely to devote more cognitive resources to learning and thus enhance performance (see also, Vanderplank, 1990).

Subtitling is an instructional aid and adding subtitles to video changes the presentation of information, thus extraneous load is expected to be generated. Therefore, in the current project, only extraneous load will be used as CL measurement. Instructional design should aim at reducing extraneous load so learners can benefit from focusing on the cognitive processing of the essential materials only. Previous literature (e.g., Kalyuga & Sweller, 2014; Mayer, 2002; Mayer & Fiorella, 2014; Mayer & Moreno, 2003; Mayer, Moreno, Boire, & Vagge, 1999) showed that information can be processed more effectively if more cognitive capacity is available. This could, for example, be achieved by avoiding the presentation of identical information through different modalities simultaneously. In relation to subtitling, Kalyuga and Sweller (2014) suggested that cognitive capacity can be increased by avoiding an addition of redundant on-screen text to material containing animation and narration in a multimedia learning environment. This is to ensure that the visual channel in working memory will not be overloaded with the presence of both animation and on-screen text and is linked to the avoidance of the redundancy effect.

2.2.2 Redundancy effect

Subtitles are part of a multimodal text with concurrent presentation of visual-verbal (written text), auditory-verbal (dialogue), visual-non-verbal (image) and auditory-non-verbal (sound) stimuli. It is inevitable that redundant information will be presented while learners process multimodal inputs in learning, especially in the case of same-language subtitles where there is a high level of redundancy between spoken dialogue and written text (subtitles), which is also known as verbal redundancy (Moreno & Mayer, 2002). It is important to understand the impact of verbal redundancy and how it affects subtitle reading and learning in order for optimal instructional procedures to be designed.

Kalyuga and Sweller (2014) explained that within CLT, the redundancy effect results when redundant information from different sources, such as visual text, diagram, graphs, pictures, speech, and music, are processed at the same time. They stated that redundant information interferes with learning rather than assisting it. When the limited working memory is designated to coordinate unnecessary information, it decreases the cognitive capacity for learning. They suggested that in order to avoid negative impact on learning as a result of the redundancy effect, redundant information that is not necessary for learning should be excluded. However, the presence of redundancy effect in audiovisual presentation is inconclusive with studies showing contradictory results. There are studies (e.g., Diao, Chandler, & Sweller, 2007; Diao & Sweller, 2007; Kalyuga, Chandler, & Sweller, 1999; Mayer, Heiser, & Lonn, 2001) that proved the presence of redundancy effect resulting in negative impact on performance, while there are others (e.g., Homer et al., 2008; Kruger et al., 2013) that showed no evidence of redundancy effect in their results. The contradictory results may suggest that the presence of redundancy effect is determined by a combination of factors within certain circumstances, or there may be alternate explanations for the processing of audiovisual materials, in which redundancy does not play a major role.

2.2.3 Modality effect: the advantage of audiovisual presentation

The term modality effect is used to describe the situation when audiovisual presentation is superior to visual only presentation (Leahy & Sweller, 2011; Low & Sweller, 2014; Moreno & Mayer, 1999). Baddeley (1992) proposed that working memory can be divided into two separate components: visual-spatial and auditory. The visual-spatial component is used to deal with information with 2- and 3-dimensions, whereas the auditory component is used to deal with spoken information including listening. If that is true, according to Leahy and Sweller (2011), there will be an increased advantage if information is processed by both processors simultaneously rather than one single processor because of the increased available capacity in working memory. When all information has to be processed solely by the visual processor, cognitive overload may be expected. If visual presentations are to be processed with both text and image, working memory capacity would reduce as a result of processing all visual information using the same channel. However, if some of the information are switched to be processed through the auditory component, learning may be enhanced. The principle of the modality effect provides some guidance for the design of effective instructions that could potentially reduce extraneous load in processing essential information. It is then logical to expect that there would be beneficial effects when information is presented in an audiovisual format (Leahy & Sweller, 2011; Low & Sweller, 2014).

Empirical research has been conducted by Ginns (2005) through a meta-analysis of 43 studies on the modality effect. Most of the studies demonstrated the effects but there have been exceptions, some studies failed to obtain the effect while some obtained a reverse modality effect. The study conducted by Tabbers, Martens, and van Merriënboer (2004) is one of the studies that obtained a reverse modality effect. Tabbers et al. (2004) demonstrated that reverse modality effect by showing that visual only presentations are superior to audiovisual presentations due to the use of lengthy verbal material, which may impose a relatively heavy working memory load when in its auditory form.

Given the definition of modality effect, the presentation of subtitles falls into the audiovisual format, and therefore may enhance learning since information will be processed by both the auditory and visual processor, and audiovisual presentation is superior to visual only presentation if no lengthy verbal stimuli are presented (Leahy & Sweller, 2011).

2.2.4 Transient information effect

Based on empirical evidence and CLT grounds, Leahy and Sweller (2011) hypothesised that the occurrence of the modality effect may be determined by how complex the auditory information is being presented, which led to the hypothesis of the transient information effect. The transient information effect occurs when written information is transformed into spoken information and result in a decrease in learning (Leahy & Sweller, 2011, 2016; Low & Sweller, 2014). Leahy and Sweller (2011) suggested that written information is a permanent format that can always be revisited and revised without having to memorise the content, and therefore does not take up extra cognitive space. Whereas the spoken information is a transient format, and that information disappears the moment it is spoken. Information needs to be memorised in order to be processed if the presentation of the auditory information is relatively complex, so extra working memory is required, therefore leaving little or no capacity for learning or comprehension. In the context of complex auditory presentation, there is difficulty in comprehending verbal information because information is not being retained long enough in the working memory.

In an audiovisual presentation, text is not the only visual stimulus, but also the presentation of image on the screen. A split attention effect could possibly occur and affect the impact of subtitles if more than one visual stimulus is presented (see also, Kalyuga et al., 1999; Mayer et al., 2001), or unless learners develop techniques and strategies in reading subtitles over a period of time to optimise its benefits in learning (Danan, 2004; Vanderplank, 1988). The nature of subtitling is different from both static written text and transient auditory information. Subtitles are semi-transient, in which text can be re-read within a short period of time, but as subtitles change according to the changes of spoken dialogues, the previous subtitle will have disappeared and it cannot be read any longer. Furthermore, the fact that subtitles interact with different sources of information creates complex viewing patterns including regressions and deflections (Romero-Fresco, 2018). It would be logical to assume that important information could be lost if the speed of subtitle presentation is too fast. The rate at which subtitles have to be displayed for optimal processing is a contentious matter. Conventionally, a subtitle speed of 12 characters per second (cps) has been used (Diaz-Cintas & Remael, 2007). A more recent study by Szarkowska and Geber-Moron (2018) showed that viewers may be able to process faster subtitles. Their research showed that viewers can read faster subtitles (20cps) without compromising their enjoyment and filmic comprehension. However, the notion of "subtitling blindness" by Romero-Fresco (2018, p. 252) describes 'a failure to notice or fully appreciate the images on the screen because attention is engaged in reading subtitles', which explains that even viewers can process faster subtitles, some visual information may be sacrificed during the delivery of the images. Lastly, the findings of Chan, Kruger and Doherty (2019) showed that fast presentation rate of verbatim subtitles (20cps and faster) generates automatically through ASR (automatic speech recognition) attenuate the benefits of subtitles in an educational context. It is obvious from the review of past research effort that questions about subtitles and its interaction with the viewers and the environment require extended investigation.

2.3 Defining subtitles and its parameters

Intralingual subtitling, also known as same language subtitles (SLS), or captioning, started appearing on television shows in the early 1970s to aid individuals with hearing impairment in understanding the information on screen (Gernsbacher, 2015; Taylor, 2005). In 1980, the term closed captioning was used when television programs begun adding captions by the National Captioning Institute so that hearing impaired individuals could gain equal access to information as part of the general public (Garza, 1991; Taylor, 2005). A closed captioned system is a special signal which can be decoded by an inexpensive decoder and makes the dialogue of the program visible on the screen as captions or subtitles[1] (Diaz-Cintas & Remael, 2007; Garza, 1991; Vanderplank, 2016). Diaz-Cintas and Remael (2007) defined subtitles as a translation practice that includes presenting a written text displaying on the lower part of the screen, that attempts to describe the dialogue of the speakers together with other elements that appear in the image, and other information contain on the soundtrack. Diaz-Cintas and Remael (2007) indicated that all subtitled programs consist of three main components: the audio text, the image and the subtitles. The basic features of the audiovisual medium are determined by the interaction of these three components, together with the viewer's ability to process both the images and the written text at a certain speed, and the size of the screen. They further emphasised that it is crucial for subtitles to appear in synchrony with the image and dialogue in order to provide an adequate semantic account of the same language dialogue, and at the same time stay on screen long enough for viewers to read them.

Jakobson (1959) identified three types of translation: intralingual, interlingual, and intersemiotic. Diaz-Cintas and Remael (2007) later classified subtitles similar to Jakobson (1959) in terms of linguistic dimensions, namely intralingual subtitles, interlingual subtitles and bilingual subtitles. Intralingual subtitles, in which spoken dialogues are transcribed into written text of the same language and displayed at the bottom of the screen; interlingual subtitles, also known as foreign language subtitles, in which spoken dialogues are translated into another language and displayed at the bottom of the screen; and bilingual subtitles, where two different languages co-exist at the bottom of the screen, where two-line subtitles are typically used with one language per line (Diaz-Cintas & Remael, 2007; Garza, 1991; Gernsbacher, 2015). Each of these linguistic parameters has their own unique functions. Intralingual subtitles are for the deaf and the hard-of-hearing, language learning, karaoke effect, dialects of the same language and for notices and announcements; interlingual subtitles are for

[1] The term subtitles will be used throughout this study to avoid ambiguity in terminology.

both hearers and deaf or hard-of-hearing users, and potentially for foreign language learners, whereas bilingual subtitles are produced to serve viewers in geographical areas where two languages are spoken (Diaz-Cintas & Remael, 2007). However, the traditional conception of intralingual subtitles as being for the deaf or hard-of-hearing users has been challenged recently and subtitles should be accessible to, and benefit all audiences regardless of their different abilities (Greco, 2018, 2019).

2.4 Contradictory findings on the effect of subtitles

The impact of subtitles is typically believed to be positive and beneficial, however, research showed differentiated results between subtitles being beneficial (e.g., Bird & Williams, 2002; Danan, 1992; Garza, 1991; Gernsbacher, 2015; Markham, 1999; Moreno & Mayer, 2002; Perego, Del Missier, Porta, & Mosconi, 2010; Vanderplank, 1988) versus subtitles being detrimental (e.g., Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Mayer et al., 2001), while some studies did not show any impact of subtitles being either beneficial or detrimental (e.g., Kruger et al., 2013; Kruger et al., 2014; Kruger & Steyn, 2013). Table 2.1 shows a summary of the effect of first- and second- / foreign language subtitles on performance.

Furthermore, results on how subtitles are processed by learners are rather inconclusive, particularly on whether subtitles result in cognitive overload. There are studies that showed the presence of subtitles creates cognitive overload (e.g., Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Mayer et al., 2001) while some of them evidenced the presence of subtitles does not necessarily create cognitive overload (e.g., Kruger et al., 2013; Kruger et al., 2014; Liao, Kruger, & Doherty, 2020; Perego et al., 2010; Vanderplank, 1988). Some studies also showed that subtitles assist form-meaning mapping (Winke, Gass, & Sydorenko, 2010), a psycholinguistic processing that connects the meaning of new vocabulary or grammar to the linguistic forms of phonological and written representations, which is essential for language learning (see also, Bird & Williams, 2002; Danan, 1992; Vanderplank, 1988; Winke, Sydorenko, & Gass, 2013).

Table 2-1

Summary of studies on the effect of first- and second- / foreign language subtitles on performance

Subtitles	Audio	Positive	Negative	No effect
L1	L1	Bird & Williams (2002) Garza (1991) Markham (1999) Moreno & Mayer (2002) Perego et al. (2007) Vanderplank (1988) Markham et al. (2001) Szarkowska et al. (2016)	Diao et al. (2007) Diao & Sweller (2007) Kalyuga et al. (1999) Mater et al. (2001)	Szarkowska et al. (2011) Guichon & McLornan (2008)
	L2/FL	d'Ydewalle & de Bruycher (2007)	None	Kruger et al. (2013) Kruger et al. (2014)
L2/FL	L1	Danan (1992)	d'Ydewalle & de Bruycher (2007)	Guichon & McLornan (2008) Bisson et al. (2014)
	L2/FL	Baltova (1999) Hayati & Mohmedi (2009) Mitterer & McQueen (2009)	None	Bisson et al. (2014) Kruger et al. (2013) Kruger et al. (2014) Kruger & Steyn (2013)

However, most of these studies are limited to language acquisition rather than content learning. When learning a second-language, the processing of the L2 subtitles is directly related to the learning goal, whereas in content learning, processing the L2 subtitles is only a tool in assisting learners to learn the actual content of that specific knowledge (van der Zee et al., 2017).

2.4.1 Positive impact of subtitles in learning: language proficiency matters

Past literature (e.g., Bird & Williams, 2002; Danan, 1992; Garza, 1991; Markham, 1999; Vanderplank, 1988; van der Zee et al., 2017) showed that language proficiency is an important factor in determining to what extent subtitles are beneficial to learners. Garza (1991) found that the use of subtitles could facilitate language use in proper context by bridging the gap between reading and listening comprehension for foreign language learners. The results showed that participants scored higher in the comprehension test and they could recall in original lexicon and collocation after watching videos with subtitles. Garza (1991) observed that the presence of the subtitles did not overload learners' comprehension ability, but rather learners with a high level of language proficiency utilised the subtitles in learning which resulted in better comprehension and retention. The presence of subtitles seems to be beneficial to advanced learners, however, it may affect learners with low language proficiency negatively in learning. Similarly, Markham (1999) also showed that language proficiency is one of the factors that affect the impact of subtitles in learning. The results showed that the availability of subtitles significantly increased the listening ability in recognising words appearing in the video content. Markham (1999) concluded that the presence of subtitles improved the listening ability in word recognition in advanced ESL (English as Second Language) students at university level.

Likewise, Vanderplank (1988) indicated that exposure to subtitles increased language learning in his study. Vanderplank (1988) conducted a nine-week longitudinal study to investigate the potential benefits of subtitle reading in language learning. The report from the participants in Vanderplank's longitudinal study indicated that habituation in subtitle reading was formed over a period of time, and that participants developed techniques and strategies in subtitle reading to maximise its benefits in language learning. Participants also reported that the presence of subtitles assisted their learning when encountering ambiguous information such as unfamiliar accents or fast-paced speech. The presentation of subtitles in the study resulted in a high level of retention and recall of language that appeared in the program content.

Vanderplank (1988) also observed that subtitle reading did not cause cognitive overload in participants, but rather it helped participants in maximising their learning by having extra processing capacity because the habituation in subtitle reading resulted in information being processed with less cognitive load, therefore more capacity was available for other processing.

Vanderplank (1988) concluded that simultaneous visual and audio presentation encouraged language learning with no evidence of cognitive overload. However, the presence of subtitles showed an opposite impact on low-intermediate English level participants, which was also evidenced in other studies (e.g., Danan, 1992; Garza, 1991; Markham, 1999) that showed language proficiency could be one of the main factors affecting the effectiveness of subtitle reading (see also, van der Zee et al., 2017). Even though a longitudinal study, the findings by Vanderplank was limited to language acquisition using materials from television programs, and it provides little information on the way L2 learners can benefit from subtitles in content learning in an educational context, which is the focus of the current study.

Moreover, the study by Danan (2004) indicated that language proficiency can affect the effectiveness of subtitles on language learning. She stated that subtitles can only be optimised when the learner has a minimum language competency threshold, and at the same time developing learning strategies to read subtitles that can maximise the benefits of subtitle reading, a similar view to the study of Vanderplank (1988). Danan (2004) further concluded that subtitles enhanced language comprehension which facilitated deeper cognitive processing for second-language learners with advance level of language competency. The presence of subtitles also seems to complement the processing of ambiguous or novel information. This perspective was supported in a study by Bird and Williams (2002), in which the result showed that the presentation of bimodal input, such as subtitles, improved the processing of audio text (spoken words) by increasing word learning and word recognition in high English level ESL learners. The result of their study concluded that dual modality (visual and audio) presentation benefits learners when information is ambiguous or novel. They explained that the process of spoken words was improved by the phonological information derived from both textual and audio presentation, and that each modality compensated the insufficiency of the other, as explained in the working memory framework proposed by Baddeley (1992, 2010, 2012).

Again, these studies were focused on the impact of subtitles on language acquisition rather than content learning.

The study by Moreno and Mayer (2002) also supported the positive impact of subtitles and at the same time demonstrated that the exposure of subtitles increased comprehension in a multimedia learning context. Their study consisted of three experiments to investigate the way the presence of subtitles promoted comprehension within a multimedia learning environment. Students were presented with non-redundant (audio only) and redundant information (audio and visual), in addition of corresponding animation and/or environmental sound. The results showed that redundant information was processed more efficiently than non-redundant information. They interpreted the result as an evidence of a dual-coding effect, in which visual and auditory stimuli can be simultaneously processed separately without causing cognitive overload (Mayer & Moreno, 1998; Mayer et al., 1999; Mayer & Sims, 1994; Paivio, 1991). They further noted that the benefits of subtitles were only limited to situations where no other visual presentations were showed to the learner concurrently. Even though the study of Moreno and Mayer (2002) focused on the impact of subtitles on multimedia learning rather than language acquisition, their study only used intralingual subtitles and provided no information on the role of linguistic format of subtitles in learning for second-language learners, which is part of the focus of the current investigation.

2.4.2 The role of language in subtitles

Most previous research showed that learners viewing L2 subtitles perform better than those viewing L1 subtitles in terms of comprehension, speech perception, listening comprehension and retention (e.g., Baltova, 1999; Hayati & Mohmedi, 2009; Lavaur & Bairstow, 2011; Mitterer & McQueen, 2009; Montero Perez, Van Den Noortgate, & Desmet, 2013), and some showed no significant impact in the presence of subtitles (e.g., Bisson et la., 2014; Kruger et al., 2014). On the other hand, Markham et al. (2001) showed that English-speaking Spanish learners reading L1 subtitles yield better performance in comprehension. The study of Markham et al. (2001) is the only study that briefly indicated the positive impact of L1 subtitles on comprehension, that is relevant to the current research. However, the main focus of their study was on language acquisition, therefore the result of reading L1 subtitles improved comprehension has not been fully addressed. Furthermore, Bisson et al. (2014) indicated that Dutch and English subtitles were read more regularly (using eye tracking measures) when video was presented in L2 audio regardless of the language of the subtitles. They explained that automatic reading behaviour (d'Ydewalle & De Bruycker, 2007; d'Ydewalle et al., 1991) may be part of the reason for the findings. The fact that both Dutch and English have similar orthographic and phonological patterns may contribute to the reading behaviour in their study. In a meta-analysis by Montero Perez et al. (2013), the result showed that L2 video with L2 subtitles outperformed the control group in listening comprehension and vocabulary learning. The meta-analysis consisted of 18 retrieved studies, among those studies 15 were included for listening comprehension analysis and 10 of those were included for vocabulary learning analysis. The findings indicated that L2 subtitles with L2 video has a positive impact on listening comprehension and vocabulary learning in the context of language acquisition. The results of these studies provide some insight into the advantage of second-language subtitles over first-language subtitles on comprehension and retention in the context of language acquisition. However, little information is given on whether L2 subtitles has the same advantage in an educational environment. It is part of the investigation in this project to find out how L1 versus L2 subtitles impact on L2 learners in the process of content learning in an academic context.

So far, most studies on subtitling primarily focus on language acquisition and the impact of intralingual subtitle on learning. The results of these studies present limited information on how second-language learners will benefit from subtitles in an educational context and the role of the language in subtitles in supporting content learning for second-language learners. Furthermore, previous research mainly focuses on the impact of subtitles on learning performance, and less on the extend subtitles impact cognitive processing or on subtitle reading behaviour. Nevertheless, some studies have been conducted to further investigate the impact of subtitles on cognitive processing using eye movement data, which will be discussed in section 2.7 of this chapter.

2.4.3 Negative impact: redundancy or individual differences?

Despite the positive results in supporting the impact of subtitles in learning, a range of different studies showed contrasting results on the impact of subtitles. There is evidence (e.g., Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Mayer et al., 2001) showing that subtitles might induce cognitive overload and thus decrease performance under different circumstances such as low language proficiency, redundancy effect and the existence of split attention effect when two visual stimuli (text and image) are presenting simultaneously. The study by Kalyuga et al. (1999) showed that duplicated information would interfere with learning instead of facilitating it under the effect of redundancy. In their study, trade apprentices and trainees were required to learn soldering on a diagram with the addition of visual on-screen text, audio text, and the combination of visual and audio text presentation. The results showed that participants scored higher in the audio text only condition, where the modality effect occurred for efficient learning. Redundancy effect occurred when visual text was presented simultaneously with the audio text, attention was split between two visual stimuli (diagram and on-screen text) and thus caused extra cognitive load resulting in decrease in performance.

Mayer et al. (2001) extended the study of Kalyuga et al. (1999) in examining the redundancy effect under a multimedia environment involving animation, subtitles, and speech in assessing retention and transfer. Their study confirmed the redundancy effect in the context of native English speakers in multimedia learning. In their study, students were required to learn from a narrated animation with or without subtitles. The results showed that performance decreased with the addition of written text presentation. They found that students' learning outcomes decreased in terms of retention and transfer when subtitles were presented. The results can be explained by the split attention occurring between different visual sources, namely animation and subtitles (Mayer & Moreno, 1998). However, students' retention and transfer outcome increased in the condition where there are no added subtitles. This can be explained by the dual coding effect where information can be processed in different modalities separately based on the assumption that the processing unit consisted of two separate channels – audio and visual (e.g., Mayer & Moreno, 1998; Mayer et al., 1999; Mayer & Sims, 1994; Paivio, 1990, 1991). Their study supported the finding of Kalyuga et al. (1999) that duplicated information presented auditory and visually in addition of the presentation of image impeded learning, and redundancy effect occurred in such circumstances. As stated by Moreno and Mayer (2002), subtitles only benefit learning when no extra visual presentation were showed simultaneously to learners, otherwise learning will be impeded. Their view is similar to the visual-textual information complexity concept proposed by van der Zee et al. (2017). That is, the amount of text presented on a video is crucial to the processing capacity of the learners. In other words, the more text being presented, the more complex the processing would become.

Similarly, Diao et al. (2007) also found the presence of redundancy effect in the context of ESL students when learning to listen a second language with the presentation of the written script. In their study, auditory materials were presented concurrently with visual text information, even though comprehension improved instantly after viewing, but performance was observed to be poorer in the succeeding auditory passage. Their study indicated that the presence of subtitles is detrimental to acquiring listening skill but effective for general comprehension and recall of content. Furthermore, their results could potentially be caused by the fact that a visual element (subtitles) was used to measure an auditory outcome (listening skill), and it was later evidenced by Sydorenko (2010) that the modality of information presentation has a positive impact on learning if the testing modality is the same. Both studies by Mayer et al. (2001) and Diao et al. (2007) indicated that redundancy effect is the factor for the decrease in performance when participants have to process duplicated information from different sources, especially with two visual stimuli presenting simultaneously, which could potentially affect learning. However, these studies still confirmed that simultaneous presentation of visual and auditory information does facilitate comprehension and recall under dual coding effect (Paivio, 1990, 1991).

Similar to the studies of Diao et al. (2007) and Mayer et al. (2001), Diao and Sweller (2007) found that subtitle reading hindered the performance of participants. In their study, a redundancy effect existed in subtitle reading for Chinese native speakers in the investigation for learning to read English as a foreign language. Their results showed that performance scores decreased and self-report on cognitive load ratings were high when information was presented in both written and audio format simultaneously. However, their study was conducted using novice learners in English as foreign language as opposed to native speaker or learners with high English proficiency. The interpretation of the results by Diao and Sweller (2007) demonstrated that language proficiency could be a main factor in the occurrence of redundancy effect.

Contrasting to the findings of Diao et al. (2007), Kalyuga et al. (1999) and Mayer et al. (2001), there are studies that showed no impact of a redundancy effect on performance (e.g., Homer et al., 2008; Kruger et al., 2013; Mayer, Lee, & Peebles, 2014). In the study conducted by Kruger et al. (2013) in the context of ESL students, they found no impact on performance with or without the presence of subtitles, but that subtitles did result in lower cognitive load. Their results suggested that if ESL students were to use subtitled video, they could experience lower CL when learning through English using same language subtitles given that they possessed a high level of English proficiency. In other words, their study showed that same language subtitles exposure could complement the learning of students in terms of facilitating information processing and understanding of the learning materials. Homer et al. (2008) also found no significant difference in learning outcome with or without subtitles. They stated that individual learning preferences – visual or verbal – together with other variables in the multimedia learning environment, such as language proficiency, language history, instructional design, presentation speed and individual reading speed, also have an impact on cognitive load.

2.4.4 Alternate explanation for contradictory findings

After reviewing some past literature, it is observed that the redundancy effect can be found in some circumstances, but not in others, and there has not been sufficient replication of these studies to demonstrate this principle conclusively either way.

Van der Zee et al. (2017) attempted to explain the contradictory findings by proposing a model through the interactions between subtitles, language proficiency and the level of visual-textual information complexity by using the attention and information processing theories. Van der Zee et al. (2017) suggested that the negative effects yielded by various studies resulted from the limited processing capacity of the working memory since both video and subtitles were sources containing visual information (see also, Kalyuga et al., 1999). They further indicated that the amount of visual-textual information of the video could hinder the effect of subtitles. Van der Zee et al. (2017) also commented on the possible effect of language proficiency as a factor in affecting the impact of subtitles as higher language proficiency required less cognitive capacity in processing the visual information.

By using their proposed model with the interaction of the three elements including subtitles, language proficiency and the level of visual-textual information complexity, the findings of van der Zee et al. (2017) evidenced that complexity of a video and the language proficiency have a significant effect in the process of learning, whereas there was no impact from the presence of subtitles. In other words, whether a learner can learn through the information presented in the video is largely dependent on the level of the learner's language proficiency and the complexity of the visual-textual information presented in the video regardless of the presence of subtitles.

Perego, Del Missier, & Straga (2018) also proposed a similar view in subtitle complexity hypothesis which provided more detail on the types of complexity that affect audiovisual processing. The hypothesis consists of three dimensions of complexity, namely structural informative complexity, linguistic complexity and narrative complexity. Their hypothesis implies that subtitles in highly complex audiovisual material could result in less effective subtitle processing and perhaps a decrease in viewing experience. The measure of structural informative complexity depends on the number of camera changes and the overall number of subtitles and its format, such as number of lines and density of information; linguistic complexity assesses the number of words and sentence length in each subtitles, and its lexical density; finally narrative complexity measures the number of primary and secondary characters, the locations where the story takes place, and the frequency of chronological alterations. The goal of their study was to investigate the effectiveness of different translation methods (subtitling vs dubbing) in cognitive and evaluative reception of different audiovisual complexity. The results showed that there was no difference in dubbing vs subtitling under the condition of moderate audiovisual material complexity. However, showing no reduction in the appreciation of the film, participants in the subtitling group showed lower cognitive performance and more effortful processing when audiovisual material is complex. The findings of Perego et al. (2018) echoed the study by Moreno and Mayer (2002) that subtitles can only be effectively processed when no extra visual presentation were showed simultaneously to viewers.

2.5 Measurement of cognitive load components

In order to understand the actual impact of subtitles in cognitive processing, it is important to measure each cognitive load component and differentiate the three different types of loads: intrinsic load, extraneous load and germane load through subjective and objective measurements (Debue & van de Leemput, 2014).

Methodologies employed in previous research include self-rating of cognitive effort, task-performance-based techniques, and physiological measurements such as eye tracking, measuring EEG (electroencephalogram) and pupil dilation (Kruger & Doherty, 2016; Paas et al., 2003). There are certain assumptions in each of the techniques that measure cognitive load. Paas et al. (2003) explained that subjective rating scale techniques are based on the assumption that people have introspection on their cognitive processes and are able to report the amount of mental effort applied. Physiological techniques assume that changes of cognitive functioning can be reflected by physiological variables, such as heart activity, brain activity, and eye activity. Task-performance-based techniques include primary task measurement and secondary task methodology. Primary task measurement is simply based on task performance, whereas secondary task methodology is based on the reflection of the cognitive load level imposed by a primary task in the presence of a secondary task. However, task-performance-based techniques can only provide information on the impact of subtitles, which does not give insight into the processing of subtitles, how and under what conditions that affects the viewers and their behaviours. Physiological techniques, on the other hand, provide an indirect measure and data on cognitive processing.

The use of eye tracking in determining the cognitive processing of information is based on the concept by Just and Carpenter (1976, 1980), who primarily proposed that the eye fixates the target information during its processing. In other words, the fixation may reflect what is being processed and the duration may reflect the time it takes to process for comprehension (Miller, 2015). This proposal is based on two assumptions, the immediacy assumption and eye-mind assumption (Just & Carpenter, 1980). In the immediacy assumption, it is assumed that readers will try to process and interpret each content word as soon as they view it and make clarification later if necessary; the eye-mind assumption assumes that the eye fixates a word as long as it is being processed. Therefore, the gaze fixation is assumed to be directly indicated by the time it takes to process a word. However, the eye-mind assumption is debatable as in reality it is not always true that information would be processed every time it is being looked at (Anderson, Bothell, & Douglass, 2004; Kliegl, Nuthmann, & Engbert, 2006). Further studies have been conducted to investigate eye movements in relation to reading pattern and information processing using eye tracking methods based on the concept proposed by Just and Carpenter (e.g., Rayner, 2009; Rayner & Reingold, 2015; Ross & Kowler, 2013).

Acknowledging that the management of cognitive load is important to optimise educational video for learning, Kruger and Doherty (2016) proposed a multimodal methodology in an attempt to measure cognitive load in the presence of subtitled educational video with precision. Their proposed methodology was based on the study by Antonenko and Niederhauser (2010), as well as previous literature review on the measurement of cognitive load. Kruger and Doherty (2016) set up an alignment of CLT constructs (including average load, overall load, instantaneous load, intrinsic load, extraneous load and germane load) and the operationalised measure for the proposed methodology including psychometric, eye tracking, and electroencephalography components. This approach allows a more extensive understanding of the dynamic nature of cognitive load, with the advantage of being able to measure a particular CLT construct at specific points, and also the impact of the redundancy effect of subtitles on learners. However, the methodology is yet to validate the uni-, bi- or multimodal context for further refinement so it can be applied to educational subtitling (Kruger & Doherty, 2016).

2.6 What is eye tracking?

Eye tracking is a technique used to record eye movements. Holmqvist et al. (2011) presented an extensive review of eye tracking measures and the history of eye trackers. According to Holmqvist et al. (2011), the earliest eye trackers, which were quite invasive, were built in the late 1800s, but only in the early 1900s eye trackers were built based on the principle of photographing external light reflection on the fovea. Different eye tracking systems may vary in their technical properties depending on its purpose, intended experiment design and context, targeted communities, and testing environment. Regardless of the range of variations, the major goal of using eye trackers is generally to record eye movements in a testing environment for research or specific domain investigation.

2.6.1 Eye movement measurements

Holmqvist et al. (2011) pointed out that fixation duration is the most reported event in eye tracking data. They stated that fixation duration is generally assumed to be the time when cognitive processing occurs during reading, and attention measurement can take place at the same time. Rayner (1998) pointed out that there is some controversy on which types of measures are the most appropriate for eye movement data potentially due to the vast possibilities in calculating eye tracking information. Similarly, Holmqvist et al. (2011) also acknowledged that the majority of measures in eye movement data contains rich spatial and temporal information and is also very versatile. They further argued that the appropriate eye tracking measures depend on what is being examined, and how experienced and familiar researchers are with eye tracking technology and measurement, and its experimental design. Despite this, there are some common measures that are generally used in reading research. Summing up from Holmqvist et al. (2011), Liversedge, Peterson and Pickering (1998), Rayner (1998), and Schotter and Rayner (2012), the main eye tracking measures used in reading research are fixation duration (the period the eye remains still in a position), mean fixation duration (MFD, the average of all fixations), first fixation duration (the duration of the first fixation on an area of interest (AOI) with single fixation or multiple fixations), first-pass reading time (the initial reading including all forward fixations), total reading time (also called dwell time or DT, the sum of all fixation durations in an AOI including saccades and regressions), regressions (eye movement back to previous words or AOI) and saccade length (also known as saccade amplitude, the distance between fixations). Schotter and Rayner (2012) indicated that fixation duration, saccade length, and percent regressions are all considered global measures, in which the average fixation duration on all words of a text will be calculated. However, if the local effects associated with fixation time on a particular word are reported, then local measures would be more appropriate than global measures. First fixation duration, single fixation duration and gaze duration are considered local measures.

In psycholinguistic research, eye tracking measurement provides ways to give objective measurements of cognitive processing directly, with the extra advantage of delivering real-time measurements (Kruger, 2016). Kruger (2016) stated that eye tracking is the only method in physiological techniques that has been used largely in subtitling research.

2.7 Eye movement in reading static text vs. dynamic text

In order to understand eye movements in subtitle reading, it is important to compare it to eye movement patterns in the reading of static text. The reading process involves a natural visual mechanism by projecting light onto the retina. The centre of the retina is the region with the highest acuity, called the fovea (Dehaene, 2009; Holmqvist et al., 2011; Schotter & Rayner, 2012). However, visual acuity decreases rapidly outside the fovea, called the parafovea, and the region beyond parafovea is called peripheral vision (Rayner, 1998). In order to effectively process information, the eyes need to make small movement, called saccades, so that the fovea can be fixated at the location where the information can be processed (Holmqvist et al., 2011;

Rayner, 1998; Schotter & Rayner, 2012). It is believed that vision is suppressed during these rapid eye movements (Rayner, 1988; Rayner, Chace, Slattery & Ashby, 2006; Schotter & Rayner, 2012). The eye movements in reading comprise of three major components, including saccades, fixations, and regression (Holmqvist et al., 2011; Rayner et al., 2006; Rayner, Juhasz & Pollatsek, 2005). Saccades are separated by pauses, which are called fixations where the eyes remain relatively still (Holmquist et al., 2011; Rayner, 1998). New information is believed to be acquired only during fixations, and each fixation typically lasts approximately 200-250ms (Rayner, 1998; Rayner et al., 2006; Rayner, Juhasz et al., 2005; Schotter & Rayner, 2012; Sun, Morita, & Stark, 1985). Fixations get longer when readers encounter difficult words or challenging content, usually a novel word or complex sentences (Rayner et al., 2006). Rayner et al. (2006) pointed out that readers may also move their eyes backwards to read the material again if the text is more difficult. This is called a regression. It is assumed that readers make longer fixations, shorter saccades, and more regressions when text is more complicated to process.

Subtitle reading is different from static reading in terms of reading speed and the interaction between different channels of information such as image, spoken dialogue, and possibly other on-screen text other than subtitles. Schotter and Rayner (2012) identified two additional tasks for subtitle viewers when reading subtitles versus reading normal text. First, subtitle viewers have to read according to the presentation time imposed by the subtitle transcribers. Since subtitles are presented within a short time period and disappear with the spoken dialogue, viewers have no control of the presentation time of the text and it is also not possible to revisit previous subtitles in the way readers do when reading static text (see also, section 2.2.4 on *Transient information effect*). Second, subtitle viewers have to switch their attention and eye location between subtitles and images (Rayner, 1988). In contrast to subtitle reading, when reading static text, readers can determine when and where to move their eyes, and they also have full control of their reading pace with only one channel of information presented at any point of time during reading.

2.7.1 Eye movement patterns in static reading: Chinese vs English

Previous studies (e.g., Feng, Miller, Shu & Zhang, 2009; Rayner, 2004; Rayner, Li, Juhasz & Yan, 2005; Schotter & Rayner, 2012; Sun & Feng, 1999; Sun et al., 1985) have found that fixation durations in reading English (270ms) are very similar to reading Chinese (260ms) despite their distinctive differences in visual form and writing system (orthography vs logography). According to Sun and Feng (1999), this similarity of eye movement patterns in reading English and Chinese suggests that fixation duration and reading eye movement patterns are determined by linguistic information rather than the visual form of the text. However, it is uncertain if eye movement patterns still remain the same when Chinese and English text are read dynamically, such as in the form of subtitles.

2.8 Eye tracking in subtitle reading

Eye tracking is an invaluable technique in the search for answers relating to the impact of subtitles and its underlying processing (see also, Doherty and Kruger, 2018). Some eye tracking studies focus on the way one- and two-line subtitles are read differently (e.g., d'Ydewalle & de Bruycher, 2007; d'Ydewalle et al., 1991; d'Ydewalle et al, 1987; Kruger & Steyn, 2013; Liao et al., 2020; Perego et at., 2010); the distribution of attention between subtitles and the rest of the screen (e.g., Baltova, 1999; d'Ydewalle & de Bruycher, 2007; d'Ydewalle et al., 1991; d'Ydewalle et al., 1987; Liao et al., 2020; Szarkowska, Krejtz, Pilipczuk, Dutka, & Kruger, 2016; Winke et al., 2013); between different subtitle language (e.g., Hefer, 2013a, b; Kruger et al., 2013; Kruger et al., 2014; Liao et al., 2020; Winke et al., 2013); and the effort required in subtitle reading (e.g., Kruger et al., 2013; Kruger et al., 2014; Kruger & Steyn, 2013). Other studies attempted to measure reading behaviour by analysing fixation duration through investigating the processing of native language and foreign language subtitles (e.g., Bisson et al., 2014; Kruger et al., 2014; Liao et al., 2020); and the impact of text chunking on subtitle processing and subtitle reading (e.g., Perego et al., 2010; Rajendran, Duchowski, Orero, Martínez, & Romero-Fresco, 2013).

An eye tracking study conducted by Specker (2015) focused on investigating the eye movements of native and non-native English speakers in subtitle reading behaviour. The results of her study showed that there is a difference in the reading pattern of dynamic text between native and non-native speakers. In the multimodal environment with eye tracking data, non-native speakers with lower language proficiency spent longer time on subtitles than native speakers. Their eye movement data showed that dynamic text was read as if it was in static condition (see also, Jensema, El Sharkawy, Danturthi, Burch, & Hsu, 2000). Furthermore, non-native speakers alternated between audio and visual modalities with the intention to look for extra information in comprehending the content. The findings also showed that the reading pattern and eye movement stay consistent for native speakers in both dynamic and static condition, whereas non-native speakers change reading pattern, indicating a change of reading strategies. The fact that language features are different in different language systems, and that essential linguistic information involves different reading strategies during language processing, explains the change in reading patterns (Koda, 1990, 1994). A later study by Winke et al. (2013) found that second-language learners spent more time in reading second-language subtitles if there is bigger distance in linguistic features between first- and second-language in terms of phonological and logographical differences. This finding is in line with the results by Guichon and McLornan (2008) and Tsai and Huang (2009) that the difference in first-language subtitles and second-language audio causes a lexical interference that may impair viewers' audio and lexical comprehension.

The findings of Specker (2015) indicated that language proficiency is a factor contributing to using text efficiently in complementing the audiovisual information (see also, Bird & Williams, 2002; Danan, 1992; Garza, 1991; Markham, 1999; Vanderplank, 1988). Furthermore, Specker (2015) stated that individual differences such as background knowledge, experience and individual learning preference for learning modalities (see also, Homer et al., 2008; Lincoln & Rademacher, 2006) also have an impact on readers' eye movements and reading patterns. The fact that non-native speakers indicated that comprehending the video content was easier with the presence of subtitles demonstrated that subtitles are still beneficial to learners with low language proficiency.

Likewise, Perego et al. (2010) conducted a study using film excerpts for cognitive processing analysis. Their study adopted a methodology of integrating the measurement of eye movement data with word recognition and scene recognition. The results showed that there was a positive effect on text and scene recognition with the exposure to subtitles in the context of subtitled film. The study also showed that incoherent syntactic segmentation in 2-line subtitles had a negative impact on information processing and recognition performance. However, the study of Perego et al. (2010) did not have a control group in the no subtitle condition, so comparison of results in performance and eye movement data is less conclusive, leaving room for further discussion.

Similarly, Kruger and Steyn (2013) used eye tracking measures to investigate subtitle reading behaviour but in the context of English subtitles on academic lectures delivered in English. Their study obtained a significant positive correlation between performance and subtitle reading, providing some evidence in supporting the usage of subtitles in education. Their study aimed to investigate subtitle reading behaviour and to formulate a way to measure the degree of reading subtitles by creating the Reading Index for Dynamic Texts (RIDT), which formed a foundation for future research on how subtitles are read and processed. Kruger and Steyn (2013) stated that the advantage of RIDT provides information that is specific for a particular subtitle being read by a particular participant with indication of the extent the subtitle was read.

They further explained that the high number of unique fixations per mean word, which is fixations excluding refixations and penalised for regression, indicated a more complete processing. However, RIDT is formulated for the use of English and would not be applicable for a language which is different in orthographic and phonological system without being validated, such as Chinese and Japanese kanji which use logogram rather than alphabet (Chang & Chen, 2002).

The study by Kruger et al. (2013) intended to gain further understanding of the impact of subtitles by measuring cognitive load using eye tracking measures, EEG and PCPD (percentage change in pupil diameter), a comprehension test and a self-report method with the presentation of subtitles. In their study, 41 Sesotho native ESL students were assigned randomly to watch lecture video in English with or without English subtitles in laboratory conditions in an educational context. Their study showed that Sesotho students reading no subtitles reported higher cognitive load, a conflicting result to most previous literature (see also, Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Mayer et al., 2001). Integrating the result of self-report frustration level and EEG data from both groups, the study indicated that participants in no subtitles group had a higher level of frustration without the support of subtitles, which suggested that the presentation of intralingual subtitles reduced cognitive load. In addition, cognitive overload was not evidenced with subtitles presentation. The study of Kruger et al. (2013) is important as it is the first empirical study to measure cognitive load directly in the presence of subtitles using EEG and PCPD, which provides important information on the processing of subtitles and its impact on cognitive load, although the use of both measures were problematic in the study. In the case of PCPD, the change of pupil diameter could have occurred due to the unstable luminosity that is not in relation to the effect of cognitive load. As for the EEG measurement of frustration level, this has not been validated for reading purposes (Kruger, 2016).

In further investigating the processing of subtitle reading, Kruger et al. (2014) conducted a study using eye tracking measures and EEG in comparing the impact of subtitle language on comprehension and attention distribution in an academic context, and the extent subtitles affect cognitive load. The current study is similar in experimental design and goal with that of Kruger et al. (2014). In their study, 68 Sesotho native ESL students were randomly assigned to watch a lecture video in English without subtitles, with English (L2) subtitles and with Sesotho (L1) subtitles. A comprehension test and self-report questionnaire on task load were used in addition to eye tracking and EEG data. Dwell time percentage and RIDT were calculated to determine the attention distribution of time spent in a certain source of information and the way subtitles were read. The results showed no significant difference in comprehension between subtitle language and the presence or absence of subtitles. However, first-language subtitles were read much less than second-language subtitles indicating that the language of subtitles have an impact on attention distribution. In addition, the results indicated that frustration levels increased with the absence of subtitles and lower comprehension effort was found when first-language subtitles were presented, similar to the result from Kruger et al. (2013) in terms of increased frustration levels in the absence of subtitles. Their results indicated that second-language subtitle reading is beneficial for short term performance, while first-language subtitle reading could result in better long-term performance given that there is more in-depth exposure.

The study of Kruger et al. (2014) supports the findings of Perego et al. (2010) on the cognitive effectiveness of subtitles while providing a more reliable cognitive load measurement with the presentation of subtitles. However, in terms of the biographical background of the participants, even though ESL students were used in the study of Kruger et al. (2014), these participants are usually much stronger in English when learning in an academic context than the Chinese ESL participants recruited in the current study. Due to the historical development of South Africa, the language of teaching and learning is mostly English regardless of students being speakers of an African native language (Kruger et al., 2014). From this perspective, it is still uncertain if the impact of subtitles will be the same on the learning of Chinese ESL students, who mostly completed their prior education in their first language, in an academic context in which teaching materials are delivered in English.

2.8.1 Subtitles and equal access to information

The work of Szarkowska focuses largely on the accessibility of information to viewers with different levels of hearing through subtitling. Her studies mainly investigate the impact of text editing and presentation rate of subtitling on comprehension and reading pattern through analysing eye movement data. Even though the work of Szarkowska is not directly related to the current project in terms of context, the results of her studies still give insight into the way hearing population behave in reading subtitles in different languages with different presentation rates.

The study by Szarkwoska, Krejtz, Klyszejiko and Wieczorek (2011) used eye movement data to investigate the impact of subtitling speed on attention distribution and comprehension on deaf, hard of hearing and hearing viewers. Their study compared verbatim, standard and edited subtitles. The results showed that viewers generally spent significantly longer time on verbatim subtitles than standard and edited subtitles regardless of their hearing status. Deaf viewers spent significantly longer time on verbatim subtitles than hearing viewers do, at the same time they also allocated more attention to verbatim than other types of subtitles. From the eye movement information, Szarkwoska et al. (2011) concluded that standard and edited subtitles are more suitable for viewing as viewers spent equal time looking at image and subtitles with no difficulty. Their study also found that deaf viewers generally preferred verbatim subtitles for complete access to information without censor. However, their study was limited by a small sample size and the difficulty of recruiting participants with similar level of hearing status from different age ranges and backgrounds.

The results of Szarkowska et al. (2016) furthered confirmed the finding of Szarkwoska et al. (2011) that deaf viewers benefit more from verbatim intralingual subtitles, and they spent significantly more time in reading intralingual subtitles because they can gain more information by lip reading as well. Their study involved Polish viewers with different hearing status: deaf, hard of hearing and hearing. Viewers were asked to watch three different genres of clips at two presentation speeds: edited subtitles at 12cps and near verbatim at 15cps. Viewers were randomly assigned to intralingual (Polish audio with Polish subtitles) and interlingual (English audio with Polish subtitles) groups. The results showed that deaf viewers obtained a higher score in intralingual than interlingual subtitles while hearing viewers performed slightly better, but not significantly better, in interlingual than intralingual subtitles. The results also showed that deaf viewers spent significantly more time on subtitles in terms of number of glances, dwell time and number of fixations. The results of the comprehension test and eye movement patterns revealed that verbatim (15cps) subtitles were read more, yielded better comprehension and displayed more effective reading patterns. Szarkwoska et al. (2016) concluded that the differences in subtitle processing between viewers in their study depends on hearing status. This finding is in line with those from Szarkowska et al. (2011) that deaf viewers prefer and benefit from verbatim subtitles.

There are other eye tracking studies that contribute distinctively to the knowledge of the processing of audiovisual translation. In the United States, the eye tracking research conducted by Jensema et al. (2000) showed that the presentation of subtitles on recorded video changed the eye movement patterns of viewers. The visibility of the added subtitles turned the viewing process into a reading process, and as the subtitle speed increased, the time spent reading subtitles also increased. The results from Jensema et al. (2000) have been supported by the eye tracking study of Romero-Fresco (2015) in the UK, which also showed that viewers allocated more time in looking at the subtitles when the subtitle speed increased, resulting in poor comprehension due to less time spent in viewing the image. The eye tracking data of the study showed that viewers distributed equal attention between subtitles and images (50%–50%) with an average subtitle speed of 150wpm (13cps). However, the distribution between subtitles and images changed to 60%-40% when subtitle speed increased to 180wpm (15cps), and 80%-20% when the speed reached 200wpm (17cps). The results on attention distribution provides further insight into the relationship between subtitle speed and subtitle reading behaviour.

2.9 Research gaps

Subtitling is an instructional aid that holds considerable potential in learning, provided that it is used in a way that will reduce the extraneous load in subtitle processing to increase learning. The language of the subtitles and its relation to the language of the dialogue and the language of the audience all impact on the effectiveness of the subtitles. Previous literature has showed the impact of subtitles mostly in the context of language acquisition (e.g., Bird & Williams, 2002; Danan, 1992; Garza, 1991; Markham, 1999; Vanderplank, 1988; van der Zee et al., 2017) and movie comprehension and appreciation (e.g., Kruger et al., 2016; Perego et al., 2010; Szarkowska & Gerber-Moron, 2018; Szarkowska et al., 2011; Szarkowska et al., 2016), but results are inconclusive in terms of whether subtitles improve or hinder performance in different circumstances. Previous results also indicated that there are a number of factors affecting the effectiveness of subtitles, including but not limited to: the complexity of visual and audio information (van der Zee et al., 2017), language proficiency (e.g., Bird & Williams, 2002; Danan, 1992; Garza, 1991; Markham, 1999; Vanderplank, 1988), redundancy of information (e.g., Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Mayer et al., 2001), cognitive complexity and presentation of information (Perego et al., 2018), modality of presentation (e.g., Ginns, 2005; Leahy & Sweller, 2011; Low & Sweller, 2014), length of subtitle exposure (Vanderplank, 1988), learners' learning preference (Homer et al., 2008; Lincoln & Rademacher, 2006) and prior knowledge and experience (Kalyuga, 2009), subtitle language (e.g., d'Ydewalle & de Bruycher, 2007; Guichon & McLornan, 2008; Kruger et al., 2014; Kruger et al., 2013; Specker, 2015), context of materials, and goals of learning. It appears that a combination of factors under certain circumstances is necessary for subtitles to be both effective and beneficial. There is a gap in the research in terms of the factors that impact on the effectiveness of L1 and L2 subtitles, particularly in an L2 educational environment. It is therefore the overarching goal of this study to investigate the factors that influence the effectiveness of L1 and L2 subtitles for second-language learners in university education. In this study, self-rating techniques, performance-based tests and eye tracking methods are used to measure the impact of subtitle reading in both objective and subjective ways, and in both real-life and laboratory condition. Different measuring methodologies are employed with an intention to form a result that is both objective and valid. Furthermore, the current study is the first in audiovisual research to combine eye tracking information from both real-life and laboratory experiments in search for the effectiveness of subtitles in an academic context.

2.10 Summary

The results in previous research on the effectiveness of audiovisual processing are inconclusive on the impact of subtitles in learning and subtitle reading behaviour. Furthermore, the investigation on the impact of subtitle reading in previous studies has not fully taken into account variables such as language history, the length of exposure to subtitles, subtitle characteristics, visual complexity and individual learning preferences. The research gap, therefore, lead to an investigation of the current study on subtitle processing by Chinese L1 and English L1 students, as well as the impact of different languages in subtitles on performance and CL with eye tracking data in the context of L2 academic environment. The goal of this research is to investigate the factors that influence the effectiveness of L1 and L2 subtitles for second-language learners in university education with three research questions, but the study will also control for the influence of other factors such as language history, exposure, characteristics of subtitles, visual complexity and individual difference.

There are two experiments in the current study. In Experiment 1, the goal is to answer Research Question 1:

> What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 or English L1 participants?

In Experiment 2, two research questions are to be answered:

> Research Question 2: What is the impact of the presence or absence of English subtitles on performance and CL for Chinese L1 participants?
>
> Research Question 3: How does the presence of English or Chinese subtitles impact on the processing of audiovisual text by Chinese L1 students?

2.11 Hypotheses

As per the review of literature in the previous sections, empirical findings from past research on subtitle reading have yielded divergent results. This is mostly due to the fact that, unlike in replication studies, most of these studies had significant variation in terms of experimental environment. This variation is evident in, among others, sample population, experimental design, types of measurement, research context, individual differences, and language use. Based on the results of past studies (Liao et al., 2020; Ross & Knowler, 2013), redundancy may not play an important role in attention allocation to subtitles in audiovisual processing. Both Liao et al. (2020) and Ross and Knowler (2013) found that subtitles were read extensively regardless of the presence of redundant information. This automatic reading of subtitles has also been established in the work of d'Ydewalle et al. (1991) and d'Ydewalle and De Bruycker

(2007). The following hypotheses are formulated based on some of the findings from past research.

Experiment 1:

Hypothesis 1.1- Chinese L1

In terms of CL, it is hypothesised that Chinese-speaking participants will have a lower CL in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles (due to the fact that the subtitles will provide written confirmation of the auditory input in English), and those reading first-language (Chinese) subtitles will have a lower CL than those reading no subtitles (due to first-language written support that may consume more cognitive resources than same language subtitles but still provide a measure of support with the benefit of being in the first-language). In terms of comprehension, it is hypothesised that Chinese-speaking participants will have a higher comprehension in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles for similar reasons to those affecting cognitive load, and those reading second-language subtitles will have a higher comprehension than those reading no subtitles. The hypotheses are shown in two simple formulas below:

Comprehension score: second-language subtitles > first-language subtitles > no subtitles

CL rating: second-language subtitles < first-language subtitles < no subtitles

The formulation of these hypotheses is based on the findings of Mayer and Moreno (1998), Mayer et al. (1999), Mayer and Sims (1991) and Moreno and Mayer (2002). The results of these studies showed that redundant information (subtitles) was processed more efficiently than non-redundant information (no subtitles) as an outcome of the dual coding effect (Paivio, 1990, 1991), in which visual and auditory stimuli can be concurrently processed separately without causing cognitive overload. The difference in CL should be pronounced in first-language

(Chinese) subtitles due to the fact that reading L1 subtitles with L2 audio may cause lexical interference (Guichon & McLornan, 2008; Winke et al., 2013).

Hypothesis 1.2 - English L1

In terms of CL, it is hypothesised that the English-speaking participants would have a lower CL in reading first-language (English) subtitles than reading no subtitles, and those reading no subtitles would have a lower CL than those reading foreign language (Chinese) subtitles. In terms of comprehension, it is hypothesised that the English-speaking participants would have a higher comprehension in reading first-language (English) subtitles than those reading no subtitles, and those reading no subtitles would have a higher comprehension than those reading foreign language (Chinese) subtitles. The hypotheses are shown in two formula below:

Comprehension score: first-language subtitles > no subtitles > foreign language subtitles

CL rating: first-language subtitles < no subtitles < foreign language subtitles

English-speaking participants, who are in the English subtitles group, would perceive lesser cognitive load in comprehension than the no subtitles group as an outcome of the dual coding effect (Paivio, 1990, 1991). However, for the other English-speaking participants, who are in the Chinese subtitles group, their perceived cognitive load may increase due to the presence of a foreign language subtitles. According to the automatic reading behaviour (d'Ydewalle & De Bruycker, 2007; d’Ydewalle et al., 1991), English-speaking participants would still allocate extensive amount of attention in reading the subtitles in a language they do not understand, which may distract them from comprehending the actual content, and therefore increase their perceived cognitive load.

Experiment 2

Hypothesis 2:

The hypotheses of the outcome of the CL and performance are the same as those in Experiment 1 (*Hypothesis 1.1*) for Chinese L1 participants.

Hypothesis 3:

It is hypothesised that participants would allocate relatively less attention in reading first-language subtitles (Chinese subtitles) than in reading second-language subtitles (English subtitles). This hypothesis is based on the findings of Guichon and McLornan (2008), Tsai and Huang (2009), and Winke et al. (2013) that lexical interference in first-language subtitles and second-language audio may impair viewers' lexical and aural comprehension, and therefore may cause avoidance in reading first-language subtitles due to this linguistic difference.

Chapter 3 Experiment 1: Methodology 1

3.1 Introduction

This chapter consists of seven sections. Section 3.2 introduces Experiment 1; section 3.3 provides information on population and sampling; section 3.4 provides information on material use in the experiment; section 3.5 describes the apparatus used in the experiment; section 3.6 elaborates on the study design and procedure, and section 3.7 provides information on data analysis including quantitative analysis on cognitive load and performance calculation and qualitative description on the eye movements. Experiment 1 (*Chapter 3*), conducted in a real-life classroom situation, aims at answering Research Question 1 (refer to section 2.10). Experiment 2 (*Chapter 5*) is a laboratory-based experiment, aims at answering Research Question 2 and Research Question 3 (refer to section 2.10). This study is a mixed-methods design involving quantitative (eye tracking measures, CL ratings and comprehension scores) and qualitative (heatmaps from eye movement patterns) analysis with subjective (self-report) and objective measurements (eye tracking and performance). The design methods are summarised in Table 3.1.

Table 3-1

Design of the current study

	Subjective	**Objective**
Qualitative	Self-report LEAP-Q	Focus maps, Heatmaps
Quantitative	Self-report CL ratings	Eye tracking, performance

Initially, the plan for Experiment 1 in the study was to use 12 lecture videos with topics aligned to the curriculum as stimuli. In this design, participants would have been required to watch two 30-minute videos every week for 6 weeks. However, the fact that the experiment was conducted in a real-world classroom meant that approval had to be obtained (and the cooperation sought) of both course administrators and teachers. This presented some problems and a compromise had to be reached, namely that the study could be conducted in class, but the duration and number of videos had to be greatly reduced to avoid disruption to the normal classroom schedule. Since only the first 10 minutes of the class time were made available for the experiment (to watch the video and complete the CL test), the only way to proceed with the experiment within the limited time frame is to reduce the 30-minute video to a 7-minute excerpt, and use the remaining three minutes for the CL test. The final experiment therefore consisted of one video viewing every week for 5 weeks, with each video being reduced to a 7-minute excerpt taken out from the middle of the 30-minute full lecture video. Participants were required to complete a 4-item self-reported CL questionnaire after each viewing, and their performance would be measured using their final examination result to avoid further classroom disruption if extra testing were administered. The interaction between CL, performance and subtitles gives an account of the way subtitles impact on audiovisual text processing. However, as the final examination also tested on content that has not been covered in the videos presented in this study, and because it was consequently difficult to isolate those parts of the examination paper that tested the content of the videos used in the study, the validity of the performance measure is arguably limited, therefore it is only used as a reference in this study. Furthermore, in the eye tracking part of this experiment it was only possible to collect four sets of eye tracking data, which means that the sample size is too small for a quantitative analysis. Therefore, these four datasets are used for qualitative analysis and to inform a subsequent experiment (Experiment 2).

3.2 Experiment 1

This was a 5-week classroom-based study with the aim to investigate the impact of subtitles on performance and cognitive load, and the possible impact of subtitle language in audiovisual text processing through a descriptive analysis of eye movements.

3.3 Population and sampling

3.3.1 Cognitive load and performance participants

A convenience sample of first-year students from a diploma program at Macquarie University was used. The diploma program is very structured and intensive with students completing two courses in a term (during a 6-week period). In addition, 75% of the students in this sample (90 out of 150), who studied in Business and Economics programs, are from a non-English background, and have to complete an introductory course on the Principles of Micro-Economics as part of their program requirement. As such, students in the diploma program constitute a representative sample of the wider population of first-year students in Business and Economics programs at the University. All students in the program were given the opportunity to participate in the experiment provided that they gave their informed consent in compliance with the ethics approval for the study (Ref. 5201700903). The sampling of this group, and the fact that the materials used in the experiment were aligned with the course curriculum, contribute to the ecological validity of the study. No specific language proficiency testing was carried out in this group mainly because of time constraints in this very compact program. However, all students have a language proficiency (IELTS) requirement for university entrance, and therefore it is reasonable to assume that the students had a comparable proficiency.

The experiment was conducted over one 6-week term (first week as briefing session and the other five weeks as experiment) of the diploma program and then repeated with a consecutive cohort in the subsequent term. In total, 16 individual classes (ten in term 1 and six in term 2) were involved in the study. Each class was randomly assigned to one of the three subtitle groups – English video without subtitles (NS), English video with English subtitles (ES) and English video with Chinese subtitles (CS). Participants who did not opt into the study with informed consent or participants with incomplete or invalid responses were excluded. Participants were required to attend all their classes for five weeks, complete five video viewings and CL ratings. There were 150 participants in Experiment 1, but data from 18 participants could not be used due to incomplete or invalid responses. The remaining 132 participants in the experiment included students from 22 countries speaking 14 different languages. Since this study focuses on English and Chinese as first- and second-language, data from 103 participants were used with 63 Chinese L1 participants and 40 English L1 participants with no knowledge of Chinese. There were 46 female and 57 male participants in this sample, and their age range was between 18 and 27. The distribution of participants in each subtitle group is showed in Table 3.2.

In order to utilise the existing data, an extra analysis including foreign language participants (non-English and non-Chinese speakers) was conducted to show how foreign language speakers behave in terms of subtitle reading when compare with those of Chinese and English speakers. The distribution of the participants including foreign language (*Others*) participants in each subtitle group is showing in Table 3.3.

In rewarding participation, only those who have fulfilled all the requirements were placed in a lucky draw for 6 prizes including one $500, one $200 and four $100 cash coupons.

Table 3-2

Distribution of participants in 3 subtitle conditions

Language	CS	ES	NS	Total
Chinese	23	26	14	63
English	15	13	12	40
Bengali	1	3	0	4
Burmese	1	0	0	1
German	0	1	1	2
Hindi	0	1	0	1
Japanese	0	0	1	1
Khmer	1	0	0	1
Korean	0	1	0	1
Nepali	1	0	1	2
Pakistani	1	2	0	3
Persian	0	1	0	1
Spanish	0	0	1	1
Vietnamese	1	5	5	11
Total	**45**	**52**	**35**	**132**

Table 3-3

Distribution of Chinese, English, and Other language speakers in 3 subtitle groups

Group	Chinese	English	Others	Total
CS	23	15	6	45
ES	26	13	14	52
NS	14	12	9	35
Total	**63**	**40**	**29**	**132**

3.3.2 Eye tracking participants

The eye tracking participants were recruited from the same population as the CL and performance measurement group. There were three criteria for recruiting participants for the eye tracking study: 1) participants' first-language must be Mandarin or Cantonese; 2) participants must be able to read and understand simplified Chinese; and 3) participants must be at least 18 years old. Five participants who fulfilled the three criteria gave informed consent to participate in the eye tracking study. However, one participant's data had to be excluded due to a low tracking ratio. Two of the remaining four participants were assigned to the condition with Chinese subtitles, and the other two to the condition with English subtitles. Participants were rewarded for their participation in the study only if they fulfilled all the requirements. These four participants were excluded in the lucky draw mentioned above, as they were part of the bigger sample from the CL and performance group.

3.4 Material

3.4.1 Videos

A series of five lectures on the Principles of Micro-Economics were used in Experiment 1. The video lectures were recorded by Dr Karunaratne (2012, 2015a, b, c, d) from the diploma program of Macquarie University. The order of the videos used in the experiment is as follows: 1) Demand & supply and government actions in markets[2]; 2) Externalities[3]; 3) Producer theory and shifting cost curves[4]; 4) Market structure and efficiency[5], and 5) Perfect competition and monopolistic competition[6]. A 7-minutes excerpt was taken from the middle of each of the video lectures to be used as stimuli in the experiment under the advice of Dr Karunaratne. In order to allow equal access of information and to avoid disadvantages in learning, the stimuli were included as part of the learned materials for all students in the program including those who did not participate in the study. Furthermore, the videos used in the study were available for the students through their iLearn (an online learning platform for students) after the experiment had been completed.

[2] Excerpt (0'0" – 6'20") can be accessed through link: https://www.youtube.com/watch?v=Z3Qb94pubAQ
[3] Excerpt (6'56" – 14'05") can be accessed through link: https://www.youtube.com/watch?v=GE8XeeaVkJ8
[4] Excerpt (1'13" – 8'24") can be accessed through link: https://www.youtube.com/watch?v=C3PDVmPijIY
[5] Excerpt (1'42"- 8'29") can be accessed through link: https://www.youtube.com/watch?v=5pxkCxMYUPU
[6] Excerpt (14'42"- 20'22") can be accessed through link: https://www.youtube.com/watch?v=3T9GSJeomKE

The videos were played through VLC media player on Window 10, which projected to a large screen inside a classroom. The videos were reduced to 75% of the original size, so the subtitles could be displayed at the bottom centre of the screen outside the video against a black frame rather than overlapping the image, as showed in Figure 3.1.

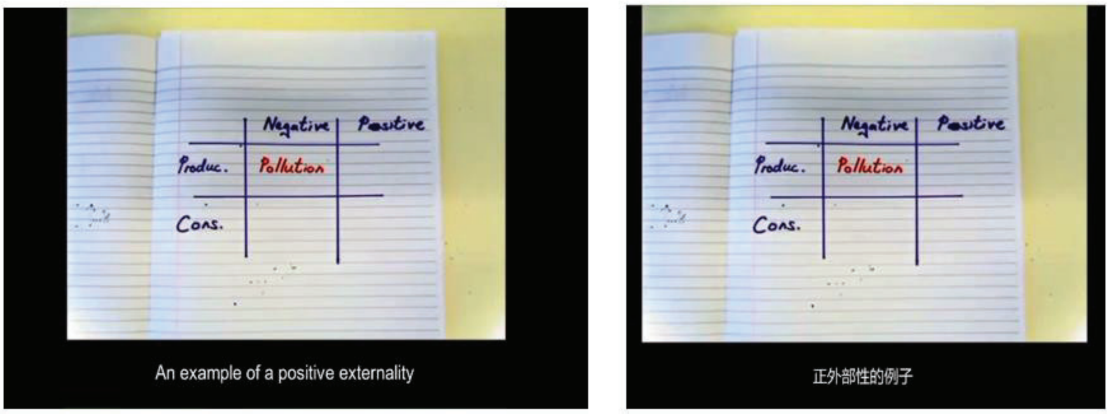


*Figure 3-1.*English (left) and Chinese (right) subtitle displays outside the video.

The style of video recording used in this experiment is paper-hand drawing style tutorial, in which the teacher talked while drawing and writing on a sheet, with only the sheet and hand visible. In addition to the subtitles at the bottom of the screen, there were graphs, formulas and text emerging on the screen as the video progressed (Figure 3.2). There is some evidence that this style of teaching presentation is a highly engaging form of online delivery (Guo, Kim, and Rubin, 2014).

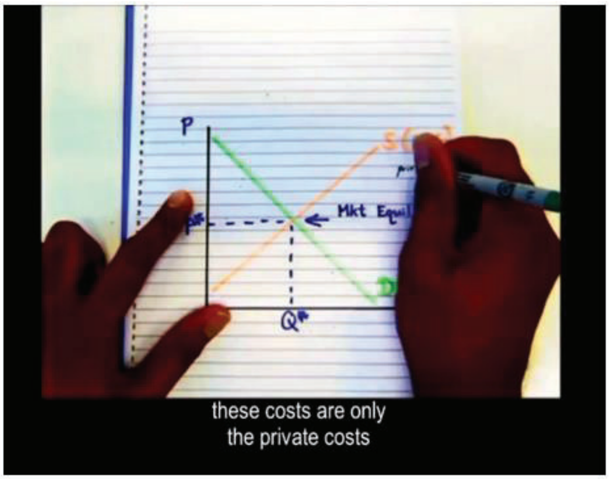


Figure 3-2. Hands of the lecturer writing on the graph.

3.4.1.1 Categorisation of redundant information

It is part of the interest of this research to investigate the factors that could impact subtitle reading. Both redundant information (Mayer et al., 2001) and the level of visual complexity of the videos (van der Zee et al., 2017) could create cognitive overload in viewers if appearing simultaneously with subtitles. In order to investigate the extent to which redundant information affects subtitle processing, the content of the five videos used in this study were coded on a continuous basis with 4 categories in terms of the presence of redundant information and hand movement of the lecturer guiding visual attention. These four categories are: 1) redundant information with hand movement (RH); 2) redundant information with no hand movement (RN); 3) non-redundant information with hand movement (NRWH); and 4) non-redundant information with no hand movement (NRN). Redundant information, in this case, refers to audio information describing the image (graphs, formulas and written text) on the screen. Sometimes, a subtitle may consist of two categories, with or without hand movement, then the subtitle will always be categorised as the one with hand movement because it would potentially provide more information on the eye movement. For example, subtitle 2.53 consists of category NRN and NRWH, viewers only see the subtitle after its first appearance (Figure 3.3 left), then the lecturer’s hands emerged with writing text or graphs towards the end of that subtitle (Figure 3.3 right).

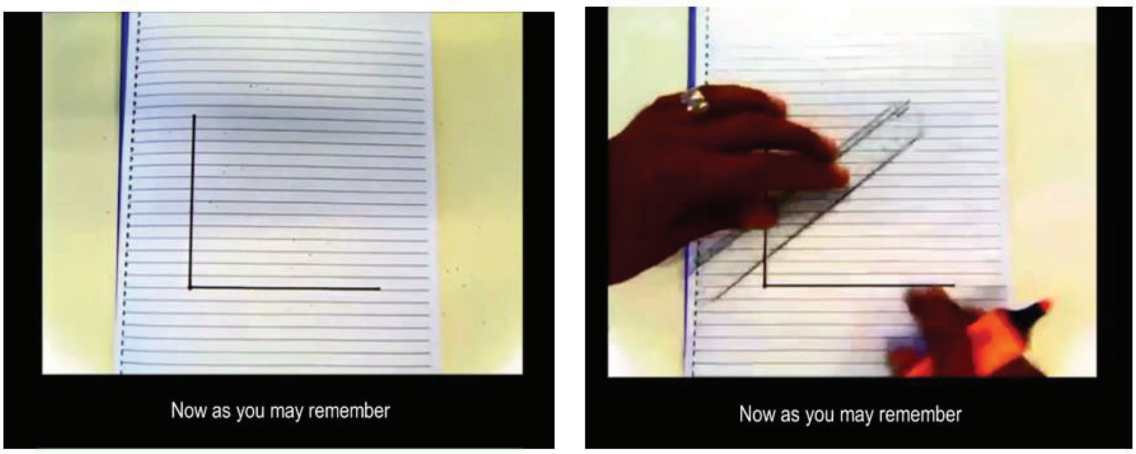


Figure 3-3. Subtitle 2.53 is categerised as non-redundant with hand movement.

In this case, the subtitle is categorised as NRWH rather than NRN because the viewers would have responded to the changes on screen by moving their eyes and thus providing more eye movement information. Figure 3.4 shows the overall distribution of the types of redundant information in these five videos.

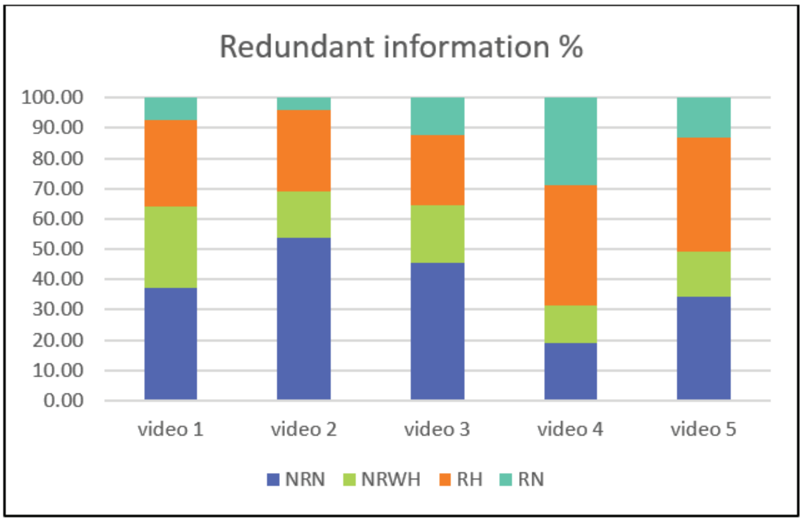


Figure 3-4. Distribution of redundant information among the five videos.

3.4.1.2 Subtitles

The videos were transcribed in English and translated into simplified Chinese respectively. The transcription and translation of the scripts were reviewed by two professional translators for accuracy. Three test conditions were created for each of the video excerpt: NS, ES, CS. Aegisub[7], a free subtitling software package that allows for the creation of professional subtitles, was used to produce all the subtitles. The display time of the subtitles from ES and CS were very close to identical[8] in order to minimise other variables that may affect subtitle reading and processing.

[7] Aegisub software can be downloaded from www.aegisub.org

[8] It is not always possible to match each subtitle in different language in terms of display time especially when English and Chinese have a very different language system. The order of the sentence could be different after translating from English to Chinese and slight adjustment in subtitling is necessary for the subtitles to make grammatical sense.

3.4.1.3 *Subtitle characteristics and quality*

All the video lectures were subtitled in English and Chinese according to established conventions (Diaz-Cintas & Remael, 2007; Ivarsson & Carroll, 1988; Kuo, 2014). The presentation rate of the subtitles adhered to subtitling conventions which is between 12-16cps for English and 1-5cps for Chinese, and the average presentation rate was around 10cps for English, and 3cps for Chinese. This is a rather slow presentation that was dictated by the slow speech rate of the lecturer, and most subtitles were verbatim transcripts. A font style of Arial in size 20 was used in the English subtitles, and Microsoft YaHet in size 30 was used in the Chinese subtitles. These fonts and size were being used for better readability. The subtitles were presented in both 1-line and 2-line format as in Figure 3.5. Previous research looked at the impact of 1-line and 2-line subtitles on subtitle processing while investigating the impact of CL on subtitle reading, but since it is not the focus of this research, the number of lines presented in each subtitles will not be taken into account during analysis.

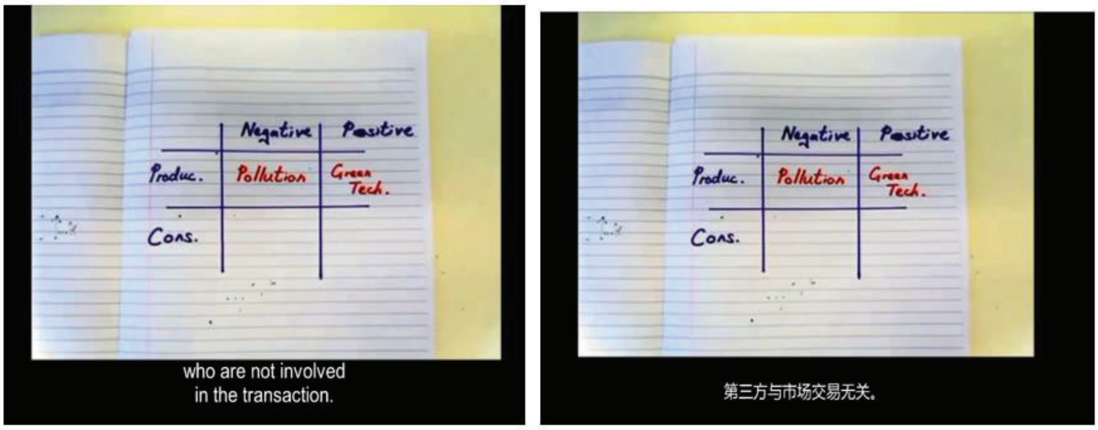


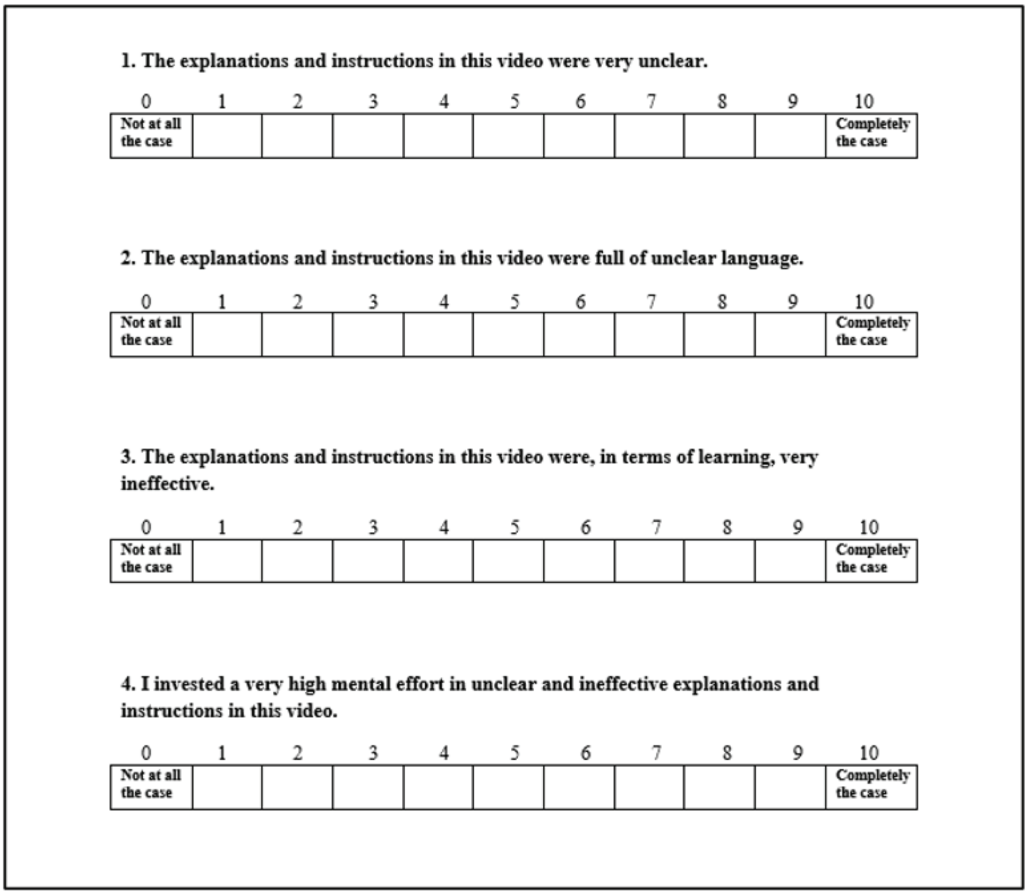
Figure 3-5. A 2-line subtitle in English (left) and a 1-line subtitle in Chinese (right).

3.4.2 Biographical questionnaire

A biographical questionnaire and LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007) were used to collect background information and language history of the participants, such as age, home language, language used in learning during secondary schooling, and years spent living in an English-speaking country (Appendix C).

3.4.3 Cognitive load instrument

A self-report cognitive load instrument adapted from Leppink and van den Heuvel (2015) was used in this experiment. Leppink, et al. (2014) developed an instrument measuring cognitive load including intrinsic load, extraneous load, and germane load respectively. The instrument was later refined by Leppink and van den Heuvel (2015) with a two-factor intrinsic/extraneous cognitive load framework, which reconceptualised germane load as a subtype of intrinsic load. Only the extraneous load instrument, which was adapted from the original instrument by Leppink and van den Heuvel (2015), was used in this study since the investigation is focused on the impact of the presentation of information. This 4-item CL instrument is a self-evaluated report on a 11-point scale from 0 to 10 with 0 being '*not at all the case'* and 10 being '*completely the case*'. The four items on the CL instrument are showed in Figure 3.6 (see also in Appendix A).



1. The explanations and instructions in this video were very unclear.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										Completely the case

2. The explanations and instructions in this video were full of unclear language.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										Completely the case

3. The explanations and instructions in this video were, in terms of learning, very ineffective.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										Completely the case

4. I invested a very high mental effort in unclear and ineffective explanations and instructions in this video.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										Completely the case

Figure 3-6. The 4-item CL self-rating questions.

3.4.4 Performance measurement

The result from the end-of-term examination was obtained with consent as the measurement of performance in Experiment 1. As mentioned in section 3.1, gaining approval for collecting data in a real-world classroom environment was itself difficult, and lecturers were not willing to approve the administration of comprehension testing in addition to the CL testing.

3.4.5 Eye tracking measurement

Dwell time and revisits were used to compare subtitle reading patterns among the four sets of eye tracking data. The subtitle reading patterns derived from the eye tracking information are presented as qualitative study.

3.4.6 LEAP-Q

Participants involved in the eye tracking study were also required to complete a LEAP-Q (Marian et al. 2007) for the purpose of interpretation in terms of language background.

3.5 Apparatus

Mobile eye tracking glasses from SensoMotoric Instruments (SMI ETG, 2016) with a sampling rate of 120 Hz was used in this experiment. The sampling rate is defined as the number of recorded samples per second (Holmqvist et al., 2011). A sampling rate of 120 Hz means there are 120 recorded samples per second. The higher the sampling rate, the more accurate the data, but a sampling rate of 120 Hz is sufficient for the type of data collected in this study. The resolution of the scene camera is 1208 x 960 pixels. The device consists of an ETG (Eye Tracking Glasses) tablet which can be attached to the SMI eye tracking glasses. Both audio and visual information can be recorded using the eye tracking device. Data were collected and analysed using SMI's iViewETG (2016) version 2.7 and BeGaze (2016) version 3.7.

3.6 Design and Procedures

3.6.1 Design

This is a 5-week study of a quantitative experimental model using three groups (two test groups and a control group), a qualitative analysis for descriptive study using focus maps from the eye tracking data, and self-reported cognitive load measures. This is a classroom-based study, which involved collecting data in a real-life classroom situation. Each group of students was randomly assigned to one of the three subtitle groups, NS (no subtitles), ES (English subtitles) and CS (Chinese subtitles). Cognitive load and performance measurement are compared between three subtitle groups, whereas eye tracking data with only two subtitle groups (ES and CS) are presented as quantitative study.

3.6.2 Procedures

3.6.2.1 Cognitive load and performance

Students enrolled in the program were given a brief introduction on the experiment by the researcher during the first lecture of their program. Consent forms and the biographical questionnaire were collected from those who chose to participate in the study and agreed to their data and information being used in the project. All students would have a chance to access the learning materials to avoid disadvantage for non-participants, however data would only be collected and used from those who gave informed consent. Since the experiment started at the beginning of each class for five weeks and would last for approximately ten minutes of the class time, participants were asked to arrive for their normal class on time to avoid data invalidation. Data would be discarded if participants arrived late.

Participants were required to watch a 7-minutes video excerpt on topics aligned to their curriculum once every week for five weeks of their program. They were then instructed to complete a 4-item CL instrument regarding the video content after each viewing.

3.6.2.2 Eye tracking measurement

Participants involved in the eye tracking study were required to wear a pair of lightweight eye tracking glasses while viewing the weekly videos in addition to the same procedure as for the CL and performance measure. They were asked to arrive ten minutes before their class time in preparation for the experiment to allow for the fitting and calibration of the eye tracking glasses. Like the other participants in the same group of the class, they were instructed to watch a 7-minutes videos excerpt on topics aligned to their curriculum once every week for five weeks of their program together with the other participants in the class. A 4-item CL instrument regarding the video content has to be completed after each video viewing. These five participants were also asked to complete a LEAP-Q (Marian et al., 2007) for their language background.

3.7 Data analysis

R (R Core Team, 2013) was used to perform all the data analysis in this study since R has the flexibility in modelling according to the nature of the data. It is an open-source statistical computing software widely used for statistical analyses in many disciplines. R (R Core Team, 2013) was initially developed by Ross Ihaka and Robert Gentleman from the University of Auckland, New Zealand. Since 1997, there has been a core group of dedicated statisticians constantly maintaining the quality of the R source, and the current R is a result of collaboration and contributions from all over the world (Crawley, 2007).

The following are the R packages that were used to analyse and plot graphs for Experiment 1: *lme4* (Bates, Maechler, Bolker, & Walker, 2015), *lmerTest* (Kuznestsova, Brockhoff, & Christensen, 2017), *readxl* (Wickham & Bryan, 2019), *ggpubr* (Kassambara, 2019), *ggplots* (Wickham, 2016), *dply* (Wickham, François, Henry, & Müller, 2019), and *car* (Fox & Weisberg, 2019).

3.7.1 Cognitive load ratings

An ANOVA was performed to find out if there is any significant difference between the three subtitle groups and two language groups in CL ratings and comprehension scores. A type III ANOVA test was used because it is designed to calculate unbalanced design with an unequal number of subjects in each group, where an interaction is present (Shaw & Mitchell-Olds, 1993) as is the case in this dataset. First-language (*Language*) was factored as an interaction with subtitle conditions (*Group*). The order of the three Group levels were factored as *no_sub*, *Chinese_sub* and *English_sub* so that the baseline contrast is a control. Tukey HSD (Tukey Honest Significant Differences) was then performed to find out which pairwise-comparison is significant between the means of groups.

An extra analysis including foreign language speaking (non-Chinese and non-English speaking) participants was also conducted for the cognitive load measurement in order to utilise the existing data. Another ANOVA type III test was conducted to find out how foreign language speakers rated CL in the three subtitle conditions. First-language (*Lang3*), namely English (n=40), Chinese (n=63) and other language (n=29), was factored as an interaction with subtitle groups (*Group*). The order of the three *Lang3* levels were factored as English, Others and Chinese so that the baseline contrast will be English because it is the same language as the audio presentation.

3.7.2 Effect size

Cohen (1988, p. 9-10) defined effect size as 'the *degree* to which the phenomenon is present in the population… The larger this value, the greater the degree to which the phenomenon under study is manifested.' In calculation, it is the difference between the means of two samples divided by the pooled standard deviation, which Cohen (1988) defined as the sum of the standard deviation of the two independent samples minus two. Cohen (1988) suggested that the effect is small if $d = 0.2$; the effect is medium if $d = 0.5$; and the effect is large if $d = 0.8$. The effect size (d) in this study is calculated through an online effect size calculator[9].

3.7.3 Comprehension scores

Kruskal-Wallis rank sum test was performed as an alternative to a one-way ANOVA test as the data was not normally distributed and the assumptions for ANOVA were not met for the test to be conducted.

3.7.4 Biographical questionnaire and LEAP-Q

Biographical questionnaire provides a brief background of all the participants and LEAP-Q provides descriptive information on the language background of the participants in the eye tracking study (Marian et al., 2007). These information may not affect the outcome of the study, but it gives extra perspectives of the background of the participants involved in the study.

3.7.5 Eye tracking data

An AOI (area of interest) was created for each subtitle in all videos. AOIs were coded according to the videos and subtitle number they belonged to. For example, subtitle number 1 in video 1 was coded as AOI 1.01, and subtitle number 50 in video 3 was coded as AOI 3.50. A reference view of all AOIs was created for gaze mapping purposes. Eye movements of the video recordings were mapped onto a reference view manually for every frame of the videos using SMI Semantic Gaze Mapping (2016) which showed the relationship between fixations and AOIs. Results were analysed and exported through BeGaze (2016) version 3.7. Total dwell time and revisits of the four eye movement data were observed and analysed. The eye tracking data in Experiment 1 formed a descriptive analysis that led to the second part of the study.

There are many possible ways to present eye tracking studies since eye tracking information contains rich spatial and temporal information (Holmqvist et al., 2011). It is impossible to

[9] Effect size calculator can be accessed through https://www.socscistatistics.com/effectsize/default3.aspx

mention every aspects of the eye tracking information in this study, but it is also not an easy task to decide what is best to include so that these qualitative descriptions can be coherently presented as part of the research. Since the qualitative part of the eye tracking study in Experiment 1 acts as a pilot study for the subsequent experiment, and the fact that only four sets of valid eye tracking data are available, it is more useful to focus on the general trend of the eye movement patterns between ES and CS group. The qualitative analysis in Experiment 1 motivates a second experiment to inform a detailed investigation on the impact of subtitle language on audiovisual text processing in a controlled laboratory environment.

For the convenience of description, the two participants reading English subtitles were coded as E1 and E2; and the two participants reading Chinese subtitles were coded as C1 and C2. All four participants were Chinese native speakers. Among a few thousand data points and numerous variables for each of these four participants in the eye tracking study, two variables are worth taking note of for the purpose of linking Experiment 1 and 2. These two variables are the total dwell time (sum of the dwell time from all participants in each group) and the total number of revisits (sum of revisits from all participants in each group). Focus maps of participant E1 and C2 on the same subtitle were compared to illustrate the difference in attention distribution when reading different language subtitles (refer to *Chapter 4* Figure 4.10). Focus maps of E1 and C2 were chosen for they distinctively demonstrated the difference in reading patterns between Chinese and English subtitles.

Chapter 4 Experiment 1: Result 1

4.1 Introduction

Results for Experiment 1 are presented in this chapter, discussion and interpretations of the results are presented in *Chapter 7*. The focus of the experiment is to investigate the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 participants. The reliability of the CL instrument and text readability of the transcripts are measured and presented, followed by a statistical analysis of CL and comprehension data through R. Finally, the eye tracking data with LEAP-Q (Marian et al., 2007) survey is presented qualitatively at the end of the chapter.

4.2 Reliability of CL instrument

Since the extraneous cognitive load self-report instrument is an adaptation from the instrument developed by Leppink et al. (2015), the contextual information of the 4-item questions has been adjusted to reflect the content of the stimuli. An analysis of internal consistency has been performed by using Cronbach's alpha coefficients to indicate between-item reliability (Table 4.1). Cronbach's alpha coefficients for the four extraneous cognitive load items is 0.899 for Experiment 1, suggesting that the items have relatively high internal consistency.

Table 4-1

Reliability test between four CL items

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N of Items
0.899	0.900	4

4.3 Text Readability

4.3.1 English transcripts

The English transcripts of the five videos were scored using the Flesch-Kincaid Readability test[10]. Four elements were tested in Flesch-Kincaid Readability test. These elements are average sentence length in words, average word length in syllables, average percentage of personal words and average percentage of personal sentences (Flesch, 1948; Kincaid, Fishburne, Rogers, & Chissom, 1975). The test score is ranged from 0-100, with 0 as the most difficult and 100 as the easiest, and a score between 60 and 80 should be quite easily comprehended by people with a school level of 7th-9th grade (Flesch, 1979). The readability scores of the transcripts from the videos are presented in Table 4.2.

Table 4-2

Flesch-Kincaid Readability score for transcripts of videos

Video number	Duration	Word number	Flesch-Kincaid Readability score (%)
1	06:20	594	54.40
2	07:10	647	60.30
3	07:10	623	74.70
4	06:48	415	50.70
5	05:40	408	78.20

As can be seen in Table 4.2, the readability scores of the video transcripts range between 50 and 80. Videos 1 and 4 are slightly more difficult, but all videos should be fairly easy to understand for the participants in this sample who are all university students.

Table 4.3 shows the overall text statistics of the five English transcripts. It can be seen that the complex word percentages are quite low in general, with video 4 having the highest percentage rate of 21%. From the result of the text analysis, the participants should be able to read the English subtitles with ease.

[10] Flesch-Kincaid Readability test can be accessed through link: https://www.webfx.com/tools/read-able/flesch-kincaid.html

Table 4-3

Text statistics from English text analysis

Video	Number of Sentence	Number of Words	Number of complex words	% complex word	Average word / sentence	Average syllables / word
1	39	594	113	19.02	15.23	1.62
2	54	647	110	16.87	12.07	1.59
3	56	623	73	11.7	11.14	1.43
4	59	415	87	20.96	7.03	1.76
5	47	408	40	9.78	8.70	1.42

4.3.2 Chinese translations

The Chinese texts were also tested for text complexity. Unlike English, there are very limited research and tools on Chinese text readability analysis due to the distinct difference between these language systems (orthography vs logography). Sung, Chang, Lin, Hsieh and Chang (2016) developed a tool called Chinese Readability Index Explorer, CRIE 3.0[11], to automate the analysis of simplified and traditional Chinese text. CRIE has the ability to analyse text in four levels of various linguistic features: words, syntax, semantics, and discourse cohesion (Sung et al., 2016). CRIE is able to perform text analysis for three types of text purpose: native Chinese speaker, Chinese Foreign Language learner, and domain knowledge. For the purpose of this study, analysis for native Chinese speaker were performed. Word difficulty can be indicated by the number of strokes in a Chinese character. Characters with 1-10 strokes is easy, 11-20 strokes is medium level of difficulty, and characters with over 21 strokes is difficult. In the Chinese text analysis result, only the elements that are equivalent to those in the English text analysis are presented in Table 4.4.

[11] CRIE online tool can be accessed through http://www.chinesereadability.net/CRIE/?LANG=CHT

Table 4-4

Chinese text analysis using CRIE 3.0

Video	Number of sentences	Number of words	Characters	Difficult word	Difficult word %	Average word / sentence	Low-stroke characters (1-10)	Low-stroke characters %
1	40	507	695	94	18.54	12.68	581	83.60
2	38	496	671	85	17.13	13.05	576	85.84
3	56	487	623	65	13.34	7.75	482	77.37
4	55	377	458	49	13.00	6.85	408	89.08
5	29	306	412	45	14.70	10.55	361	87.62

It can be seen from Table 4.4 that the difficult word percentage are quite low in general, with the highest percent rate of 18.54% for video 1. The low-stroke characters percentage are very high, between 75% to 90%, which indicates that the Chinese texts are also very easy to read. From the result of the text analysis, the Chinese subtitles are fairly easy for the participants as well. The full analysis of the Chinese text can be seen in Appendix E.

The following section is to present the data in relation to the specified research questions.

4.4 What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 participants?

4.4.1 CL results

The ANOVA summary for CL (see Figure 4.1 of Appendix D) showed that there is a significant ($F(2, 397) = 6.55$, $p < 0.05$, $\eta p2 = 0.03$) interaction between *Group* and *Language*. As the ANOVA test is significant in CL, a Tukey HSD test can be performed to find out the multiple pairwise-comparisons between the means of groups. The formula and its output summary are presented in Figure 4.2 of Appendix D. The results showed that two pairwise-comparisons are significant with an adjusted p-value < 0.05. CL between English speakers reading ES (first-language subtitles; $M = 15$, $SD = 9.43$) and English speakers reading CS (second-language subtitles; $M = 9.47$, $SD = 7.36$) is statistically significant ($t\ (138) = 5.49$, $p < 0.05$, $d = 0.65$);

CL between English speakers reading ES (first-language subtitles; *M* = 15, *SD* = 9.34) and Chinese speakers reading ES (second-language subtitles; *M* = 10.2, *SD* = 7.27) is statistically significant (t (193) = 4.72, $p < 0.05$, $d = 0.57$). The boxplot in Figure 4.3 shows the mean CL ratings in the three subtitle groups, and there is an interaction between *Group* and *Language*. English-speaking participants reading CS rated significantly lower in CL than those reading ES; and Chinese-speaking participants reading ES rated significantly lower in CL than English-speaking participants reading ES.

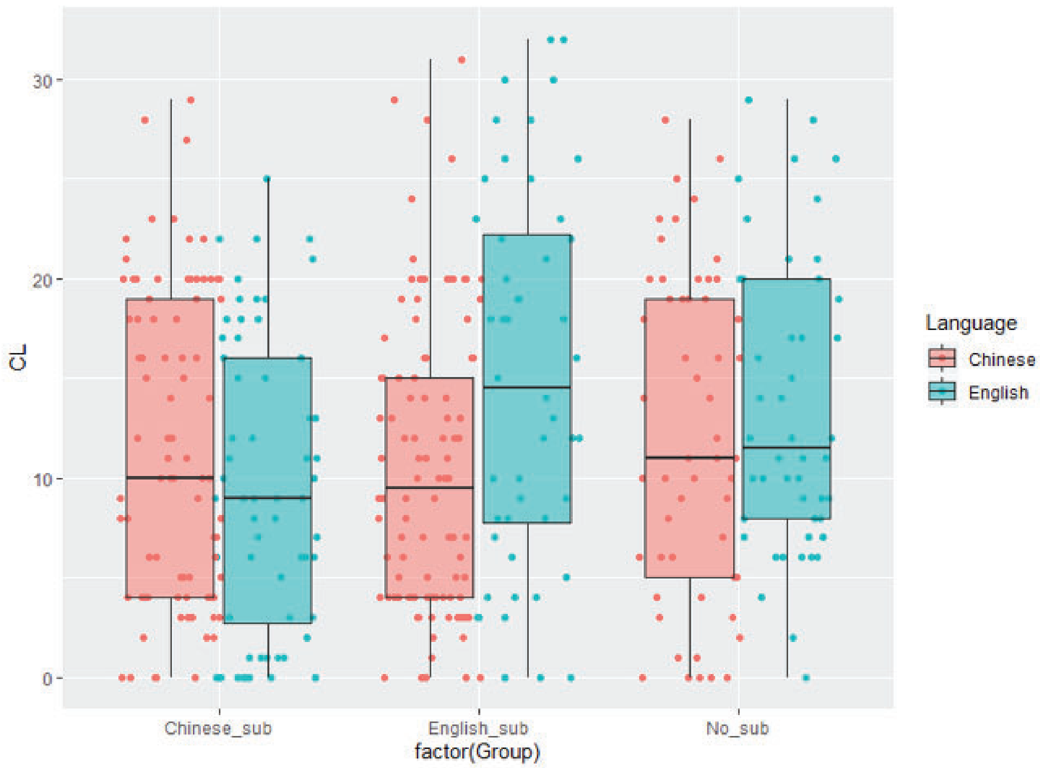


Figure 4-3. CL ratings between Chinese and English speakers.

4.4.2 Adding other languages into the analysis

The formula for ANOVA and its output summary in R is presented in Figure 4.4 in Appendix D. The results showed that the difference in CL ratings between ES (*M* = 8.86, *SD* = 8.82), CS (*M* = 8.37, *SD* = 8.33) and NS (*M* = 9.31, *SD* = 8.34) is statistically significant (F (2, 651) = 3.95, $p < 0.05$, ηp2 = 0.012), and the difference in the interaction between *Group* and *Lang3* is also statistically significant ($F(4, 651) = 3.20$, $p < 0.05$, $\eta p2 = 0.019$). The boxplot in Figure 4.5 shows the mean CL ratings between three language groups, and there is an interaction between *Group* and *Lang3*. The results showed that foreign language participants reading no subtitles reported significantly lower CL ratings than those reading Chinese (foreign language) subtitles and English (second-language) subtitles, and those reading English (second-language) subtitles reported lower CL ratings than participants reading Chinese (foreign language) subtitles, making CL rating the highest in the CS group.

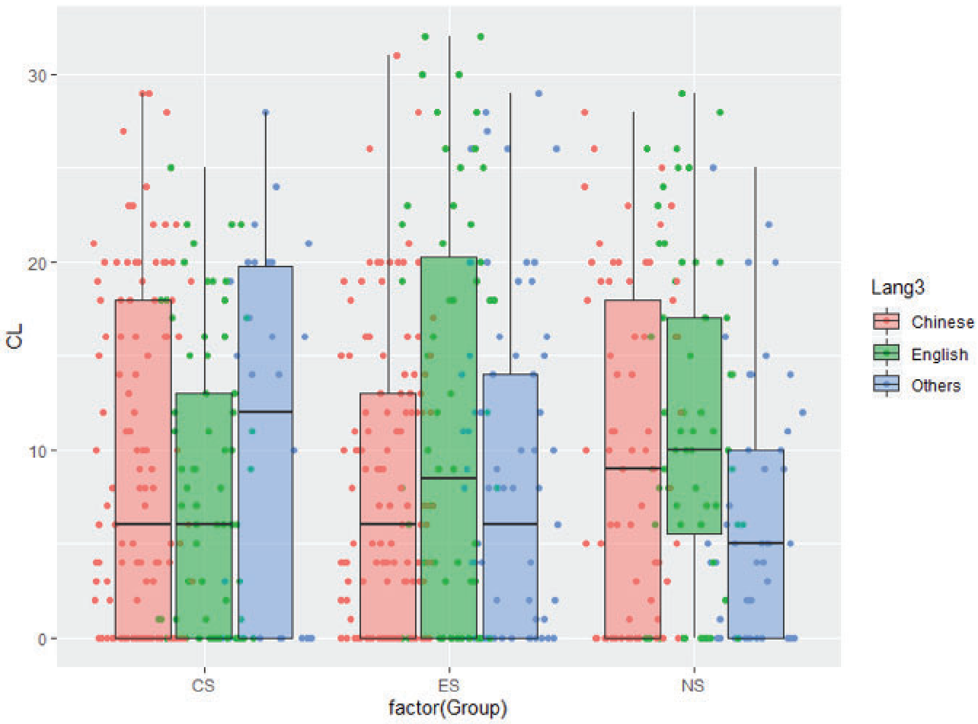


Figure 4-5. Average CL ratings between Chinese vs English vs other language speakers.

4.4.3 Comprehension result

The Kruskal-Wallis rank sum output summary is presented in Figure 4.6 in Appendix D. The output showed that the difference in comprehension between the Chinese speakers ($M = 35.9$, $SD = 9.12$) and the English speakers ($M = 30.9$, $SD = 8.87$) is statistically significant (X^2 (1, N = 103) = 7.48, $p < 0.05$, $d = 0.56$); and the difference in comprehension between the three subtitle groups (NS: $M = 34.8$, $SD = 9.1$; CS: $M = 32.2$, $SD = 9.97$; ES: $M = 35.1$, $SD = 8.75$)

is not statistically significant (X^2 (2, N = 103) = 1.69, $p > 0.05$). Since there were only two levels for *Language* (English and Chinese), no further test was needed as the difference has already been proven significant.

The boxplot in Figure 4.7 shows the mean comprehension scores in these three subtitle groups. The Chinese-speaking participants scored significantly higher in comprehension than the English-speaking participants in all subtitle groups (X^2 (1, N = 103) = 7.48, $p < 0.05$, $d = 0.56$); and there is no significant difference in comprehension between subtitled conditions (X^2 (2, N = 103) = 1.69, $p > 0.05$). However, it is worth noting that the validity of the comprehension scores from Experiment 1 is questionable because results were taken from the end of term exam which included content that were not covered in the videos.

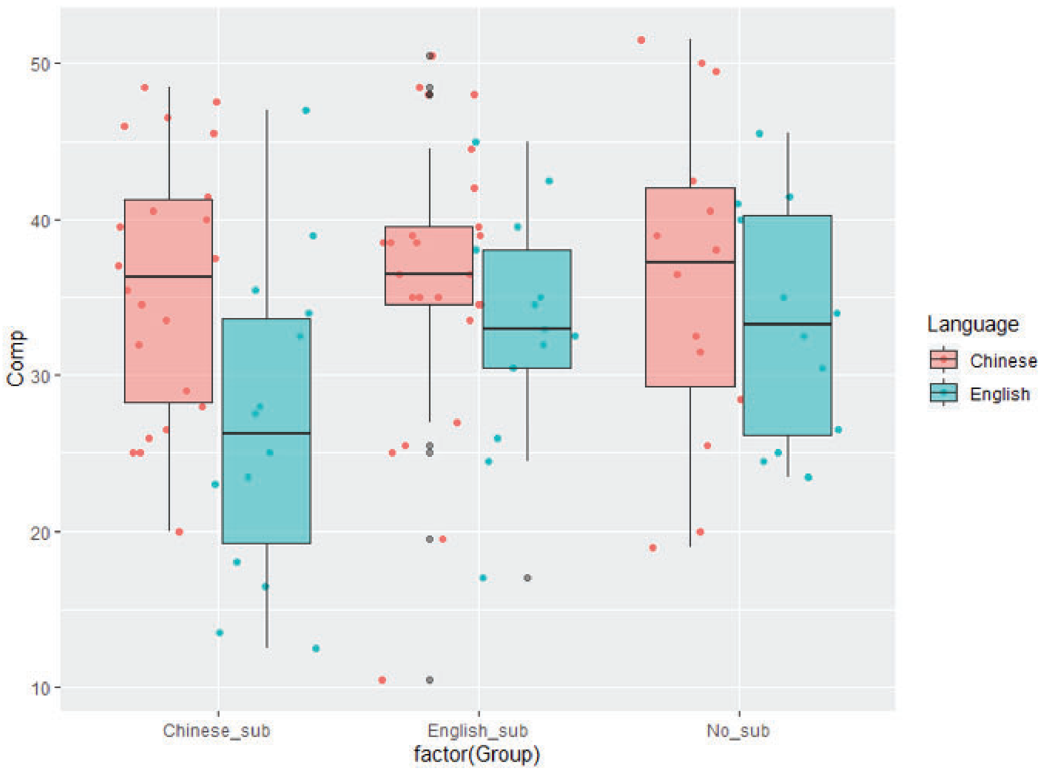


Figure 4-7. Average of comprehension scores between NS vs CS vs ES.

4.4.4 Biographic questionnaire and LEAP-Q

The biographic questionnaire provides a brief background of the 103 participants involved in the CL and comprehension task including gender, age, home country, secondary schooling language, and time living in English-speaking countries. There were 63 Chinese-speaking participants and 40 English-speaking participants (46 female and 57 male) in Experiment 1. The participants were aged between 18 and 27, with an average age of 19.6 years old. Among the 63 Chinese participants, 46 used Chinese and 16 used English as the learning medium in secondary school, and the average time spent in English speaking countries was 24 months, with the shortest time being less than one month and the longest time being 84 months. The biographic data therefore shows that the background of these Chinese participants in Experiment 1 in terms of secondary schooling language and time living in English-speaking countries varies. It is, however, showed in the analysis that the biographical factors have no statistically significant impact on the CL test and comprehension, therefore the detail statistical computation is not presented here.

The LEAP-Q (Marian et al., 2007) of the four eye tracking participants showed that even though there were individual differences, the language background of the participants was otherwise very similar in general. All of them started learning English between ages 10–15 and indicated that Chinese is their dominant language. They have been living in an English speaking country for an average of 6 years (with the shortest as 1.5 and the longest as 11 years) at the time of the experiment, and they do not speak English with family with the exception of one participant. They reported more exposure to Chinese than English, which was expected, and they learned English through friends and media such as radio and music. Furthermore, they perceived their level of English proficiency as slightly more than adequate in terms of speaking, reading, and understanding.

4.4.5 Qualitative study on eye movements

Total dwell time and total number of revisits of the four eye tracking Chinese-speaking participants were plotted in Figure 4.8 and Figure 4.9 according to their subtitle group. Figure

4.8 shows that the total dwell time between participants reading English and Chinese subtitles are quite similar. In other words, participants allocated similar attention in their first- and second-language subtitles.

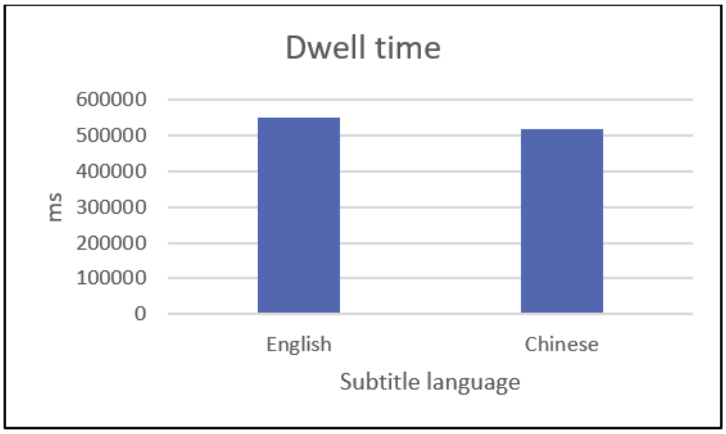


Figure 4-8. Total dwell time difference between two subtitle language groups.

However, the number of revisits in the Chinese subtitles group were more than double those in the English subtitles group, as is showed in Figure 4.9. The graphs indicate that participants who saw English subtitles had fewer revisits than those who saw Chinese subtitles. On the other hand, more revisits with similar dwell time means participants who saw Chinese subtitles switched between subtitles and images more frequently than those reading English subtitles.

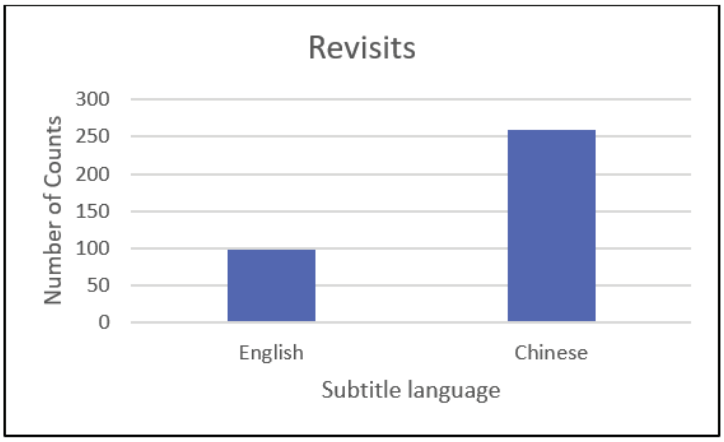


Figure 4-9. Total number of revisits difference between two subtitle language groups.

A focus map comparison between participant E1 and C2 on the same subtitle is presented in Figure 4.10. It illustrates that Chinese speakers reading first-language subtitles had a more interrupted reading pattern and switched their attention between subtitles and screen regularly, whereas Chinese speakers reading second-language subtitles had a more uninterrupted reading pattern on the subtitle.

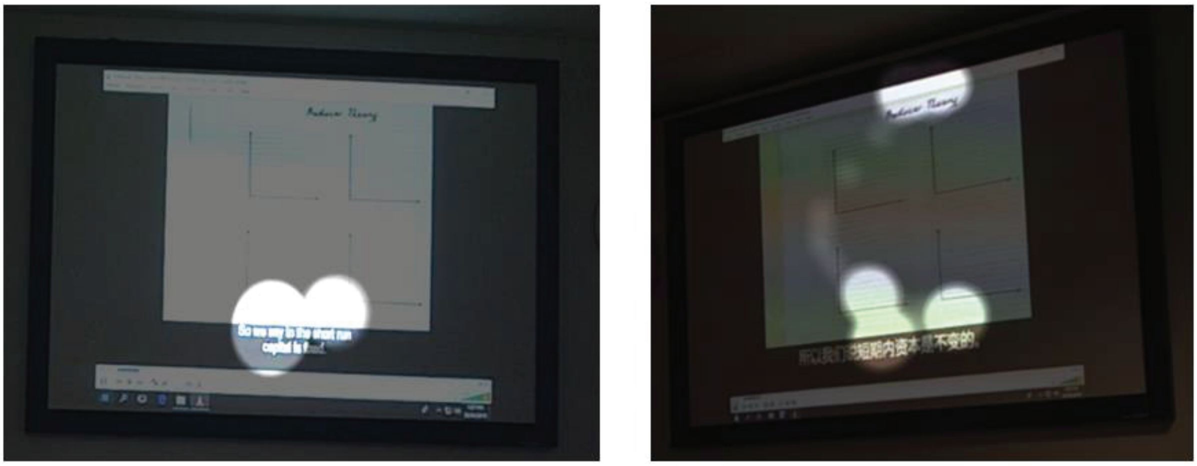


Figure 4-10. Focus maps of participant E1 and C2 reading the same subtitle in English (left) and Chinese (right).

4.5 Summary of results

The results of Experiment 1 showed that English-speaking participants reading second-language subtitles rated significantly lower in CL than those reading first-language subtitles; and Chinese-speaking participants reading second-language subtitles rated significantly lower in CL than English-speaking participants reading first-language subtitles. For the Chinese participants, the eye tracking data seems to indicate that they may have been more inclined to switch more regularly between image and subtitles and that this more fragmented reading may reflect less efficient cognitive processing in the presence of the Chinese subtitles, although this cannot be confirmed by the limited eye tracking sample. The comparison in eye movement patterns of Chinese speakers between Experiment 1 and 2 will be further discussed in *Chapter 7* section 7.6. In view of the limited validity of the comprehension scores mentioned in *Chapter*

3 Introduction and further discussed in *Chapter 8*, the comprehension results in Experiment 1 are not particularly informative, although it will be discussed briefly. Chinese-speaking participants scored significantly higher in comprehension than English-speaking participants in all subtitle groups, and there is no significant difference in comprehension between the three subtitle groups regardless of the first language of the participants. The brief qualitative analysis of the four eye tracking data sets suggests that the language of subtitles has an impact on the way participants process audiovisual text. Further investigation on the impact of subtitle language in audiovisual text processing is presented in *Chapter 5* and *Chapter 6*.

Chapter 5 Experiment 2: Methodology 2

5.1 Introduction

This chapter presents the methodology of the second experiment across seven sections. Section 5.2 introduces Experiment 2; section 5.3 provides information on population and sampling; section 5.4 provides information on material use in the experiment; section 5.5 describes the apparatus used in the experiment; section 5.6 elaborates on the study design and procedure, and section 5.7 provides information on data analysis. Due to the acknowledged limitations of Experiment 1, the design of Experiment 2 aims at answering the question that could not be answered in Experiment 1 regarding to the role of language in subtitle processing.

The findings from the four sets of eye tracking data in Experiment 1 indicated that subtitle language has an impact on the processing of audiovisual text. It has been observed that with similar attention distribution on subtitles, Chinese speakers reading second language subtitles tended to read along with the second-language audio whereas those reading first-language subtitles showed more interrupted reading patterns by frequently switching between subtitles and images. However, the limited sample size of this exploratory eye tracking experiment did not allow for any reliable conclusions on the role of subtitle language in processing audiovisual text and its impact on CL and performance. Therefore, a second experiment was designed and conducted in a more controlled environment to investigate the impact of subtitle language on subtitle processing and comprehension.

5.2 Experiment 2

This experiment involved a remote eye tracking system (SMI RED250, 2011) to analyse the processing of L1 and L2 subtitles while also measuring self-rated cognitive load and comprehension in a laboratory environment. The material used in Experiment 2 are the same as those in Experiment 1, except for the measurement of comprehension as will be explained below.

5.3 Population and sampling

Seventy participants with Chinese (Mandarin or Cantonese) as their first-language were recruited for this experiment. Four criteria were used for recruiting: 1) participants' first-language must be Chinese, 2) participants must be able to read and understand simplified Chinese, 3) participants must be at least 18 years old, and 4) participants must be studying postgraduate or above degree or have a postgraduate qualification. Participants (57 female and 13 male) aged between 22 and 55 were recruited from Macquarie University. Since Experiment 1 involved students in a pathway diploma with a wide range of proficiency, the second experiment was designed to have more control over the language proficiency, recruiting students in a postgraduate course to guarantee a higher language proficiency. Ethics was approved (Ref. 5201700903) and individual informed consent was obtained in conducting the experiment. All participants were rewarded a payment of $40 cash for their time spent in the experiment.

5.4 Material

5.4.1 Videos

The videos used in Experiment 2 were the same as those in Experiment 1 (see, *Chapter 3,* section 3.4.1).

5.4.1.1 Subtitles

The same subtitles were used in Experiment 2 as in Experiment 1. Since Experiment 2 was conducted through SMI Experiment Centre (2010) version 3.7 in SMI RED250 (2011), complex procedures were involved to process the videos and subtitles to the acceptable format before they could be imported to the software. First, the auditory quality of the videos was adjusted by Audacity[12], an open source software, due to the inconsistent volume level among the five videos. A separate mp4 audio file was extracted from each video using Wondershare Video Converter Ultimate[13], then the volume of these audio files was adjusted by the amplified function in Audacity. These adjusted audio files were saved as wave files. The new wave files were then combined with the original videos, replacing the original soundtrack through Windows Live Movie Maker[14]. Since Experiment Center (2010) does not accept separate subtitle files (such as srt and ass files), the subtitles had to be hardcoded to the videos before it could be imported into Experiment Center (2010). The subtitles were hardcoded at the bottom center of the videos and were then cropped into the 7-minutes excerpts used in the experiment through Wondershare Video Converter Ultimate. Unlike in Experiment 1, subtitles in Experiment 2 can only be hardcoded onto the videos and therefore inevitably overlapping the image due to technical limitation. A font style of Arial in size 12 was used in the English subtitles, and Microsoft YaHet in size 12 was used in the Chinese subtitles. Figure 5.1 shows the ways subtitles are displayed in Experiment 1 and Experiment 2.

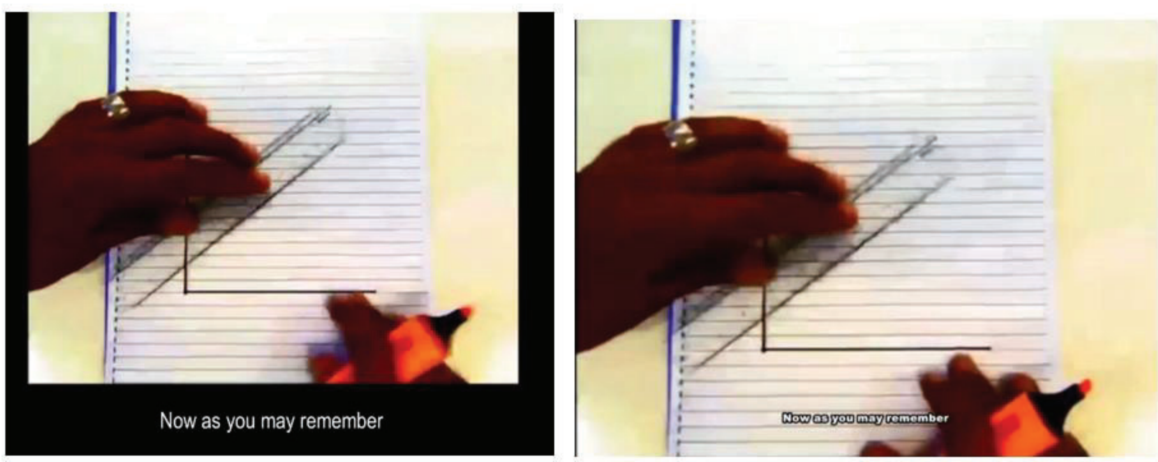


Figure 5-1. Subtitles display differently in Experiment 1 (left) and Experiment 2 (right).

[12] Audacity software can be downloaded through: https://www.audacityteam.org/download/
[13] Wondershare Video Converter Ultimate can be downloaded and purchased through https://videoconverter.wondershare.com/
[14] Windows Live Movie Makers can be downloaded through: https://windowsprores.com/windows-live-movie-maker/

There is an obvious difference in the size of the subtitles between Experiment 1 and Experiment 2 because of the different environment these two experiments were conducted in. Experiment 1 was conducted in a classroom situation where the videos were projected onto a large screen, therefore a much larger font size (see, *Chapter 3*, section 3.1.4.3) was used to ensure the subtitles were readable for the participants. On the other hand, Experiment 2 was conducted in a laboratory where subtitles were presented through a 17 inches laptop screen, therefore the subtitling size is relatively smaller than those in Experiment 1.

5.4.2 Questionnaires

The same biographical questionnaire and 4-item CL instrument were used in Experiment 2 as in Experiment 1 (Appendix C and A).

5.4.3 Performance measurement

A comprehension task based on the content of the five videos was created for the participants to complete immediately after the CL instrument. Due to time limitation and avoid fatigue in the participants, only five multiple choice questions were used for each video, and a total of 25 comprehension questions. A different comprehension task was created in Experiment 2 because of the fact that the end-of-term examination in Experiment 1 did not test enough knowledge linked to the five videos. The comprehension questions used in this experiment were sourced from online sample questions by Frasca (2007) as well as questions created by the researcher based on the video content (Appendix B).

5.5 Apparatus

SMI RED250 (2011), and software iViewX (2016) and Experiment Centre (2010) version 3.7 were used in Experiment 2 to collect eye tracking data. The device has a sampling rate of 250 Hz and a screen resolution of 1920 x 1080 pixels. The stimuli were showed on a 17 inches laptop screen. The experiment was developed and recorded using iViewX (2016) and

Experiment Center (2010) version 3.7; and BeGaze (2016) version 3.7 was used to code and analyse the eye movement data.

5.5.1 Eye tracking ratio

In this experiment, eye movement data with tracking ratio less than 85% were removed from analysis resulting in 39% loss of the original data. Tracking ratio can still be very low despite a good calibration. Holmqvist et al. (2011) suggested a list of possible factors that could influence the precision of eye movement measurement, and some of those factors apply here as well. The reasons behind the unpredictable tracking ratio could both be physical and behavioural. Physically speaking, if a person has drop eyelids, it could be more difficult for their eyes to be detected by the eye tracker because part of the eye was covered by the eyelids. Secondly, if a person has eye problems such as astigmatism or is severely short-sighted, then the eye tracking could potentially be problematic. Thirdly, the height of a participant could affect the accuracy of measuring eye movement as it changes the angle between the eyes and the eye tracking device. In terms of behaviour, different people will have different ways of watching video or looking at the computer screen. Some of these ways could potentially make tracking the eye movement difficult if the person looks at the screen with the head up or down, forcing more of their eyes to be either covered by the eyelid or blocked by the eye lashes. Eye measurement accuracy reduces in both cases. Furthermore, eye tracking accuracy will be compromised if a participant is wearing heavy eye makeup, coloured contact lenses or shaded optical glasses. Motivation also affects the behaviour of the participants. Participants who tended to continuously move their body parts (legs, hands, and head) are the ones who found the content of the videos difficult to comprehend or lack the interest to engage fully. These body movements eventually affect the accuracy of their eye movements being measured. Some of these mentioned issues are easier to overcome than others, for instance, a participant reminder can resolve the problem of heavy eye makeup and coloured lenses; a verbal reminder during the experiment could also help participants to be more conscious and alert that their body movement may affect their eyes being detected properly. However, other issues such as drop eyelids, the level of vision clarity and motivation are harder to control once the participant is already in the process of doing the experiment. Potentially, a more specific recruitment criteria could exclude those with severe visual problems; and a topic that is more interesting and less technical could be used to resolve the motivation issue.

5.6 Design and Procedures

5.6.1 Design

This experiment took approximately one hour to 90 minutes to complete per participant with a quantitative experimental model using three-group (two test groups and a control group) design. This is a laboratory-based study involving the collection of data in a controlled environment. The participants were randomly assigned to one of the three groups in the experiment, no subtitles (*n*=18), English subtitles (*n*=28) and Chinese subtitles (*n*=24) respectively (see Table 5.1). Cognitive load, comprehension scores, and eye tracking data such as fixation count (FC), mean fixation duration (MFD), dwell time percentage (DT%), time to first fixation (TTFF) and revisits were calculated and compared between these three test groups. The detail description of these measures can be found in section 5.7.5 to 5.7.9 of this chapter.

Table 5-1

Distribution of participants in each subtitle condition

Group	Chinese_sub	English_sub	No_sub	Total
Chinese speakers	24	28	18	70

5.6.2 Procedure

Each participant completed the experiment individually inside a laboratory with the SMI RED250 (2011) eye tracking device. Experiment Center (2010) version 3.7 and iViewX (2016) were used to collect data. Participants were asked not to wear heavy eye make-up and coloured contact lenses to the experiment. They were briefed about the experimental procedure, then were asked to sign the consent form before the experiment started.

Participants were asked to sit comfortably in front of the eye tracking device which was attached to a laptop. Their positions were adjusted within the range of the device, and participants were asked to maintain a similar position throughout the experiment for their eye movement to be recorded successfully. The laboratory was sufficiently illuminated for eye tracking purpose. Participants were asked to switch off their phones to avoid distraction, and there was no noise from outside the laboratory after the door was closed, since the experiment was conducted in a sound-insulated laboratory.

At the beginning of the experiment, a 9-point calibration and 4-point validation was conducted to ensure participants' eyes were calibrated with the device for accurate eye tracking measurement. Participants were then required to complete a biographical questionnaire on the laptop before the experiment started. Full instructions on the experimental procedure was given on the screen as the experiment proceeded. The first video excerpt started after a successful calibration followed by the completion of the biographical questionnaire. Participants were required to complete a 4-item cognitive load instrument and answer five multiple-choice comprehension questions after each video viewing. There was no time limit for participants to answer the questions. A 30-second break was given after the comprehension and before the next calibration. After the short break, another calibration and validation started again before the next video, and the process repeated for all five videos. The whole experiment lasted approximately one to one and a half hour depending on the speed of each participant in responding to the questions, and if there was any issue during calibration. The study procedure is presented as a flow chat in Figure 5.2.

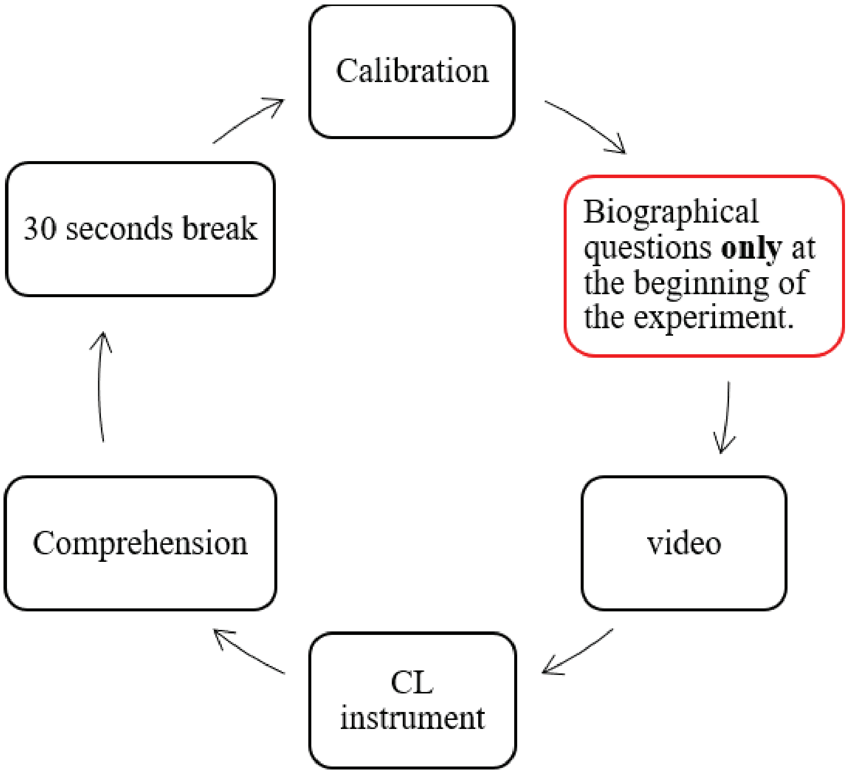


Figure 5-2. Study procedure in Experiment 2.

5.7 Data preparation and analysis

R (R Core Team, 2013) was also used to perform the analyses in Experiment 2. The following are the R packages that were used to analyse and plot graphs for this experiment: *lme4* (Bates et al., 2015), *lmerTest* (Kuznestsova et al., 2017), *readxl* (Wickham & Bryan, 2019), *ggpubr* (Kassambara, 2019), *ggplots* (Wickham, 2016), *dply* (Wickham et al., 2019), *pROC* (Robin et al., 2011), *glmmTMB* (Brooks et al., 2017) and *DHARMa* (Hartig, 2019).

5.7.1 Cognitive load ratings

An ANOVA type III test was performed due to the unequal number of participants in each subtitle group. The aim is to find out if there is any significant difference between three subtitle groups and two language groups in CL ratings.

5.7.2 Comprehension scores

In terms of comprehension, since the data was not normally distributed, a Kruskal-Wallis rank sum test was performed to calculate the difference between subtitle and language group in comprehension scores as an alternative to a one-way ANOVA test. A Wilcoxon pairwise-comparison test was then performed to find out which pairwise-comparison is significant ($p < 0.05$). Average comprehension scores and average cognitive load ratings were compared between the five videos.

5.7.3 Eye tracking data

AOIs were created around each individual subtitle to include all eye movement data (fixations and saccades) during the reading of subtitles. These eye movement data was then exported per AOI for analysis with an emphasis on fixation count (FC), mean fixation duration (MFD), dwell time percentage (DT% - the percentage of time the participant looked at the subtitle as percentage of the time the subtitle was on screen, also referred to as proportional reading time), time to first fixation (TTFF) and revisits. A Generalised Linear Mixed Model (*GLMM*) was used to analyse the variables. Detail statistical computations and R output scripts can be seen in Appendix D.

Firstly, the dataset needed to be scaled and cleaned before they can be calculated as showed in Figure 5.3 of Appendix D. A subset was created using 99% quantile to discard the outliers without discarding too much of the data. An extra variable *processed* was then created to distinguish the zero and non-zero data in the FC column. In other words, an argument was created by setting all FC data that were larger than zero as subtitle being processed. Following that, *CPS* (character per second) and *MW* (mean word) were scaled and centred. This is a very important step to form a logical analysis with a successful outcome. During the computation of *glmer* involving *CPS* and *MW*, the intercept is where *CPS* and *MW* equals zero, and zero *CPS* and *MW* means there is no subtitle which does not make logical sense. Therefore, *CPS* was centred as the average speed for each subtitle language so that the intercept was not zero *CPS*. It is not advisable to use the average *CPS* for both languages combined since character has a different meaning in English and Chinese. Character in English comprises a single alphabet whereas character in Chinese comprises a morpheme (Schotter & Rayner, 2012). The average word length for Chinese is 1.5 character (Sun et al., 1985) and 5 characters for English (Bochkarev, Shevlyakova, & Solovyev, 2015), therefore the speed for Chinese subtitles is much slower than those in English subtitles. The outcome will not be accurate if the analysis is based on the *CPS* that average both languages combined. However, by centring *CPS* at the average for each language individually, the outcome of the analysis would be more accurate when the language difference is being accounted for. The same applies to *MW*, which needs to be centred to the average *MW* in order to make the analysis outcome logically accurate. However, unlike *CPS*, *MW* can be scaled just for the average of both languages combined rather than individually because the average word length for English and Chinese has already been normalised and are equivalent to each other (Schotter & Rayner, 2012; Sun et al., 1985).

5.7.3.1 Binomial modelling

A *GLMM* binomial model was used followed to calculate the probability ratio of the subtitles being processed or not processed. The first step of the analysis was to get rid of the outliers by scaling and cleaning the data, and 99% quantile was used to discard the outliers that could potentially skew the outcome. After that, a *ROC* (receiver operating characteristic) curve and a boxplot were used to visualise the fitting of the model. A *ROC* plot shows the performance of the proportion of positive (sensitivity) and negative (specificity) observations that are correctly classified as the output, which could be above or below the threshold, and produces a finite set of data within the *ROC* space (Fawcett, 2006; Robin et al.,2011). A correlation between subtitle language, skipped subtitles, redundancy, CL ratings and comprehension scores were calculated.

5.7.3.2 Fixation counts (FC)

Number of fixations was used to indicate various reading features including semantic importance, search efficiency and difficulty, and word properties in reading (Holmqvist et al., 2011). By calculating the number of fixations in the presence of other factors such as redundant information, subtitle language and subtitling speed, some light is shed on the factors that could affect reading in first- and second-language subtitles. It is necessary not to interpret fixation count alone but to incorporate the results from mean fixation duration and dwell time percentage as well because the number of fixations is a very general measure. Fixation count alone only provides information on the frequency of fixations without indicating the length of each fixation. The relationship between dwell time and number of fixations becomes obvious when several short fixations with shorter dwell time in one AOI is compared to a smaller number of long fixations with longer dwell time in another AOI (Holmqvist et al, 2011).

A *GLMM* analysis was calculated after the dataset had been prepared (see Figure 5.3 in Appendix D), and 99% quantile was used to discard the outliers. The order of the four *Redundant* levels were factored as NRN (non-redundant with no hand movement), NRWH (non-redundant with hand movement), RH (redundant with hand movement) and RN (redundant with no hand movement) so that the baseline contrast was a control (non-redundant with no hand movement); the order of the two *Group* levels were factored as *Chinese_sub* and *English_sub* so that the baseline contrast was Chinese subtitles as it was the first-language of the participants. Five models have been considered for model comparison. These five models factored in redundant information *Redundant*, subtitling speed *CPS_centred*, and mean word of each subtitle *MW_centred* as fixed effects (+) to or an interaction (*) with the subtitle language *Group*. *DHARMa* (Hartig, 2019) in R package was used to test the homogeneity and normality of the residual of the best fitted model through simulation. The *DHARMa* package uses “a simulation-based approach to create readily interpretable scaled residuals for fitted (generalised) linear mixed models" (Hartig, 2019, p.1). The scaled residuals are created by simulating from the fitted model through the *simulateResiduals* function with the number of simulations set to n=1000. Hartig (2019) explained that the function has three purposes, firstly, a new artificial dataset will be created through simulating from the fitted model; secondly, a cumulative distribution of simulated values will be calculated for each observed value; and finally, a quantile value that corresponds to the observed value will be provided. If the specified model is correct, it is expected to observe a uniform distribution of the overall residuals and a uniformity in y direction when the residuals are plotted against any predictor (Hartig, 2019).

5.7.3.3 Mean fixation duration (MFD)

As previously described in *Chapter 2* section 2.7, fixation duration is one of the most commonly used eye tracking measurement in researching eye movements, and it is defined as the time duration when the eye is relatively still (Holmqvist et al., 2011). It is generally assumed that fixation duration is the time when information acquisition occurs during reading (Holmqvist et al., 2011) and new information is being processed (Rayner et al., 2006). Since each subtitle may consist of more than one fixation with different fixation durations, an average of the fixation durations of a subtitle was used in the measurement. By calculating and comparing the mean fixation duration between the two subtitle groups, we can observe the ways these subtitles were being processed.

A *GLMM* was used in this analysis. As with the other eye movement variables, the dataset needed to be prepared as in Figure 5.3 in Appendix D, using 95% quantile to discard the outliers. As the normal range of mean fixation duration is between 50ms and 600ms during reading (Rayner, 2009), the value of the mean fixation duration was 594ms at 95% quantile of the data, which was within the generally acceptable range (with 99% being 960ms). The order of the four *Redundant* levels were factored as NRN, NRWH, RH and RN so that the baseline contrast was a control (NRN, non-redundant with no hand movement), and the order of the two

Group levels were factored as *Chinese_sub* and *English_sub* so that the baseline contrast was Chinese subtitles as it was the first-language of the participants. Seven models were considered for model comparison. Residuals were tested and graphed, and boxplots were used to compare the average of the mean fixation duration between subtitle groups.

5.7.3.4 Dwell time percentage (DT%)

Dwell time is defined as the "sum of durations from all fixations and saccades that hit the AOI" (SMI, 2009, p.368). Holmqvist et al. (2011) suggested that dwell time on an object could indicate an interest from a participant, or an object consists of useful information, however a longer dwell time may be an indication of uncertainty, poorer situation awareness, and difficulty in extracting information from a display. Dwell time percentage was used in this analysis rather than dwell time alone because dwell time percentage is a better measurement in comparing between participants by showing the proportion of time spent on each subtitle. Dwell time percentage is calculated as the sum of the duration of all fixations and saccades that hit the AOI (dwell time of the AOI) divided by the duration of the AOI (SMI, 2009). Incorporating results from dwell time percentage, mean fixation duration and fixation counts together when interpreting the outcome could provide a better perspective on the impact of subtitle language in audiovisual text processing.

Generalized Linear Mixed Models using Template Model Builder (*glmmTMB*) (Brooks et al., 2017) was used for this analysis. Brooks et al. (2017) indicated that many types of *GLMMs* including count data and continuously distributed response can be fitted by *glmmTMB*. The benefits of choosing *glmmTMB* is its ability to fit zero-inflated mixed models (the current dataset displays a lot of zero responses) with speed and flexibility, and has a similar interface to *lme4* at the same time (Brooks et al., 2017). According to Brooks et al. (2017, p.378), one of the unique features of *glmmTMB* among other packages that can fit zero-inflated mixed models is "its ability to estimate the Conway-Maxwell-Poisson distribution parameterized by the mean”, but Gamma distribution was used in the current analysis. Same as fixation counts, the dataset needed to be prepared as in Figure 5.3 of Appendix D, using 99% quantile to discard the outliers. The order of the four *Redundant* levels were factored as NRN, NRWH, RH and RN so that the baseline contrast was a control (non-redundant with no hand movement) and the order of the two *Group* levels were factored as *Chinese_sub* and *English_sub* so that the baseline contrast was Chinese subtitles as it was the first language of the participants. Twelve models were considered for model comparison. Residuals were tested and graphed. Stacked bars and boxplot were used to compare dwell time percentage between subtitle groups.

5.7.3.5 Time to first fixation (TTFF)

Time to first fixation can be defined as the duration of the time until the first fixation hits the AOI after the onset of the AOI. There is no direct extraction from the software for this measurement, and it can only be calculated by subtracting ‘Time to First Appearance’ from ‘Entry time’. ‘Time to First Appearance’ is defined as the “time when the AOI becomes visible for the first time relative to the trial start” (SMI, 2019, p.367); and ‘Entry Time’ is defined as the “duration from start of trial to the first hit of AOI” (SMI, 2019, p.368). Time to first fixation provides a latency measurement on the first fixation of the AOI after it becomes visible for the first time. It may take longer for a participant to make the first fixation on the AOI (in this case the subtitles) after its first appearance if there are distractions (graphs, written text and hand movement) on the screen and this may also reveal whether there is a difference between processing L1 and L2 subtitles.

GLMM was used to analyse time to first fixation. Data was scaled at 99% quantile to discard the outliers, and scaling for *CPS* and *MW* is not necessary since they are not included in the time to first fixation analysis based on a reasonable assumption that subtitling speed have no impact on the time to first fixation. The order of the two *Group* levels were factored as

Chinese_sub and *English_sub* so that the baseline contrast was Chinese subtitles as it was the first language of the participants. Three models were considered for comparison.

5.7.3.6 Revisits

Revisit is also termed as returns, refixations or rechecks by different researchers and eye tracking manufacturers (Holmqvist et al., 2011). The number of revisits to an AOI always equals the number of glances in the AOI minus one as there has to be at least one glance to have a revisit. Holmqvist et al. (2011) indicated that revisits occur when the areas are semantically informative.

Initially a *glmer* was considered to analyse the number of revisits, however, none of the models can fit the data properly and the residuals have not been tested successfully. In order to have a better look at the data, two graphs were plotted to examine the raw data from the ES and CS group for *CPS* and *CPS_centred* at 99% quantile.

5.7.3.7 Effect size

The effect size (d) in this study was calculated through an online effect size calculator[15], as per Experiment 1, with $d = 0.2$ as a small effect, $d = 0.5$ for medium effect, and $d = 0.8$ as a large effect.

5.7.3.8 Subtitle processing plots for ES and CS

Five subtitle processing graphs were plotted to visualise the way English and Chinese subtitles were being processed. Individual subtitles were tagged in the subtitle processing graphs to indicate when CS and ES showed 100% or close to 100% processing rate (yellow triangle), when CS and ES showed more than 25% difference in processing rate (red diamond), and when CS and ES showed low processing rate (orange circle).

[15] Effect size calculator can be accessed through https://www.socscistatistics.com/effectsize/default3.aspx

In *Chapter 6*, section 6.5.4, only a small number of heatmaps from a few representative AOIs are presented to visualise the eye movement patterns and the circumstance where the participants distributed their attention on the screen. This section presents a few heatmap comparisons from a number of chosen AOIs to provide a visualisation on the participants' attention distribution on the screen in the presence of English or Chinese subtitles.

Chapter 6 Experiment 2: Result 2

6.1 Introduction

Results for Experiment 2 are presented in this chapter, discussion and interpretations of the results are presented in *Chapter 7.* The focus of the experiment is on investigating the role of subtitle language in audiovisual text processing. Research question 2 and Research question 3 are to be answered: What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 participants? How do the presence of English and Chinese subtitles impact on the processing of audiovisual text by Chinese L1 participants? The reliability of CL instrument and text readability of the transcripts are measured and presented, followed by statistical analyses of CL ratings, comprehension scores and eye tracking data. Fixation counts, mean fixation duration, dwell time percentage, time to first fixation and revisits are the eye tracking measures used in answering the second research question. Eye movement patterns are presented visually through graphs and heatmaps as descriptive analysis in the last section of this chapter.

6.2 Reliability of cognitive load instrument

An analysis of internal consistency has been performed for item reliability of the same CL instrument as in Experiment 1 by using Cronbach's alpha coefficients (Table 6.1). Cronbach's alpha coefficients for the four extraneous cognitive load items is 0.868 for Experiment 2, which is very similar to the coefficients obtained in Experiment 1 (0.899; see, *Chapter 4*, Table 4.1), suggesting that the items have relatively high internal consistency.

6.3 Text Readability

The readability scores for both English and Chinese text are the same as those in Experiment 1 (Table 4.2 – 4.4 in *Chapter 4*) since the same transcripts were used in both experiments.

The following section is to present the data in relation to the specified research questions.

Table 6-1

Reliability test between 4 CL items

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N of Items
0.868	0.874	4

6.4 What is the impact of the presence or absence of English or Chinese subtitles on CL and performance for Chinese L1 participants?

6.4.1 CL measurement

Secondary schooling language (*SL*) has been considered as a factor in the computation, but since it had no impact on CL ratings (see Figure 6.1 in Appendix D), *SL* will not be included in the ANOVA analysis.

The ANOVA summary for CL in Figure 6.2 of Appendix D showed that the difference between Group (NS: $M = 10.7$, $SD = 8.93$; CS: $M = 10.1$, $SD = 8.33$; ES: $M = 9.79$, $SD = 8.70$) is not statistically significant ($F(2, 347) = 0.28$, $p > 0.05$, $\eta p2 = 0.002$). The boxplot in Figure 6.3 shows the CL ratings between the three test groups.

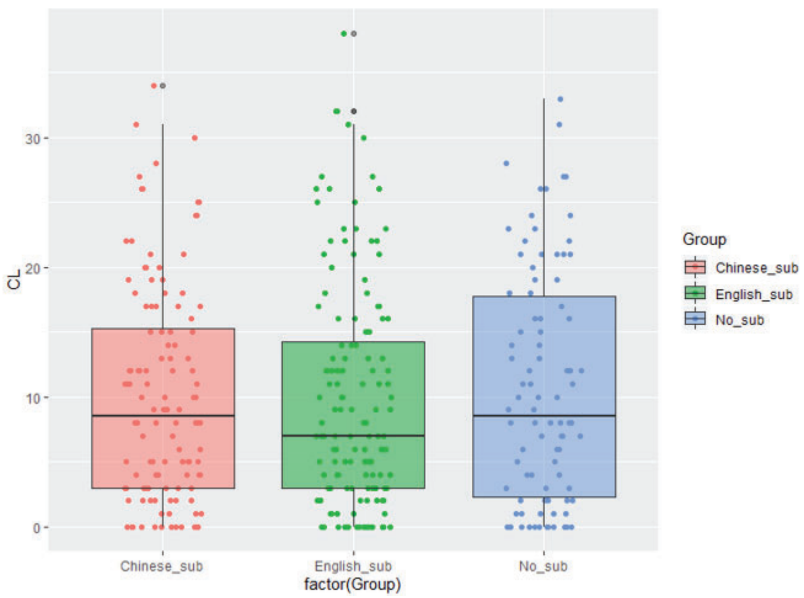


Figure 6-3. CL ratings between CS vs ES vs NS.

6.4.2 Comprehension measurement

The Kruskal-Wallis rank sum test summary in Figure 6.4 of Appendix D showed that the difference in comprehension between these three test groups (NS: $M = 3.13$, $SD = 1.21$; CS: $M = 3.52$, $SD = 1.02$; ES: $M = 3.19$, $SD = 1.12$) is statistically significant (X^2 (2, N = 70) = 1.69, $p < 0.05$). As the test is statistically significant in comprehension, Wilcoxon pairwise-comparison test can be performed to find out the multiple pairwise-comparison between the means of groups.

The Wilcoxon test output in Figure 6.4 of Appendix D shows that 2 pairwise-comparisons are statistically significant with an adjusted p-value < 0.05 (red box). The comprehension between CS ($M = 3.52$, $SD = 1.02$) and NS ($M = 3.13$, $SD = 1.21$) is statistically significant ($p < 0.05$, $d = 0.35$); the comprehension between CS ($M = 3.52$, $SD = 1.02$) and ES ($M = 3.19$, $SD = 1.12$) is also statistically significant ($p < 0.05$, $d = 0.31$). Chinese participants scored significantly higher in viewing Chinese subtitles than viewing English and no subtitles; and there is no significant difference in comprehension between viewing English and no subtitles (ES: $M = 3.19$, $SD = 1.12$; NS: $M = 3.13$, $SD = 1.21$; $p > 0.05$). Raw data can be found in Appendix F. The boxplot in Figure 6.5 shows the comprehension scores in these three test groups.

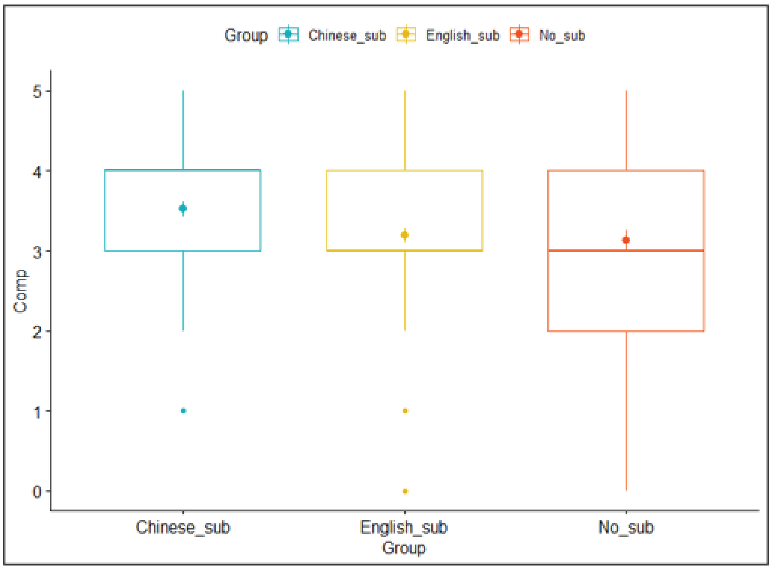


Figure 6-5. Mean of comprehension scores between CS vs ES vs NS.

6.5 How do the presence of English and Chinese subtitles impact on the processing of audiovisual text by Chinese L1 students?

6.5.1 Eye tracking analysis: Binomial modelling

A Generalised Linear Mixed Model (*GLMM*) was used for the eye tracking analysis. A binomial model was used, and it focused on the odds ratio of the subtitles being processed or not processed based on the language of subtitles (*Group*) as fixed effects; the participants (1|Participant) and each subtitle per video (1|video/AOI) as random effects. The binomial modelling is based on the concept of probability, and the odds ratio (also known as likelihood) are defined as "the ratio of the probability of an event A occurring divided by the probability of an event not occurring" (Winters, 2017, The odds ratio, para. 2). In this case, the modelling is to calculate the ratio of the probability of English subtitles being processed to the probability of Chinese subtitles being processed.

A boxplot was used to examine model fit on a *GLMM*. Figure 6.6 indicates the probability of processed and non-processed subtitles being predicted as processed and not processed by the model.

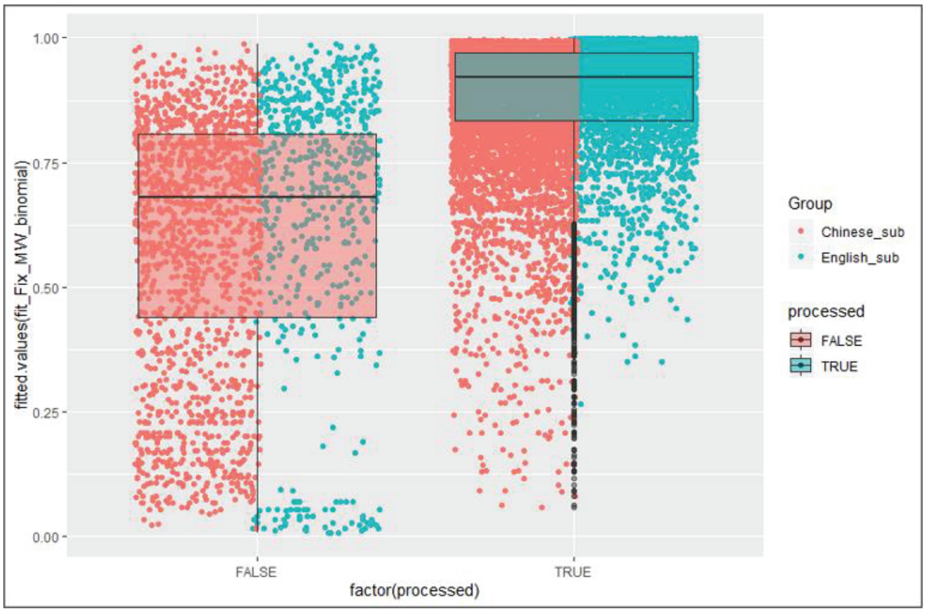


Figure 6-6. TRUE indicates the probability in predicting processed subtitles being processed is 0.90.

In the boxplot, TRUE indicates that the median probability of processed subtitles being processed is approximately 0.9, and FALSE indicates that the median probability of non-processed subtitles not being processed is approximately 0.6. The boxplot shows that the model has a high accuracy in predicting correctly that the processed subtitles would be processed. This is evidenced in Table 6.2 in the actual percentage of subtitles being processed (83%) and skipped (17%) in raw data with 99% quantile. Therefore, it can be concluded that this model fits well in the data and has an accurate prediction of the result.

Table 6-2

The odds ratio for processed and non-processed subtitles in raw data with 99% quantile

Language	# Subtitles processed	# Subtitles skipped	Grand Total
Chinese	5414	1503	6917
English	5270	631	5901
Grand Total	10684	2134	12818
Total percentage	83.35%	16.65%	100%

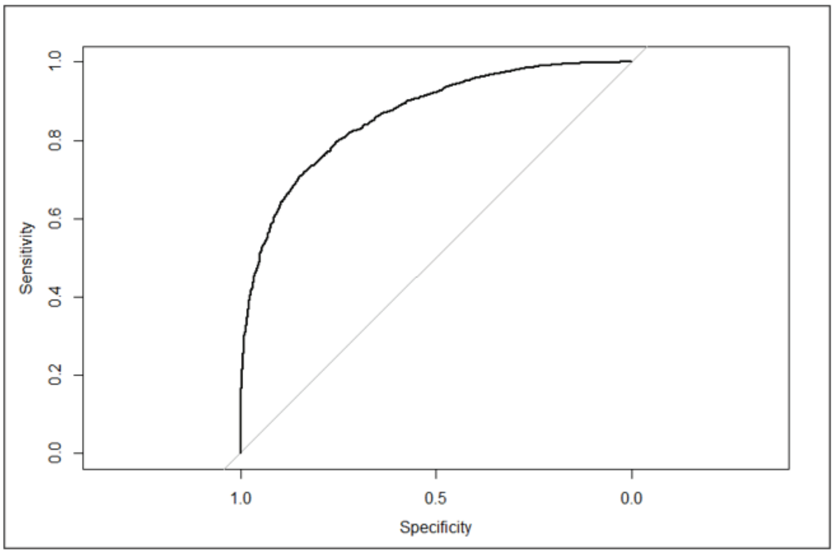


Figure 6-8. The area under curve (AUC) is 0.86.

A *ROC* curve (Robin et al., 2011) was plotted to visualise the model fit as showed in Figure 6.8. The area under the curve (AUC) measures the performance of an observation of a classifier in the *ROC* context. A higher AUC that is closer to 1 indicates a good model fit (Robin et al., 2011). The area under the current *ROC* curve is 0.85 (see Figure 6.7 in Appendix D), which indicates the current binomial model is a good model fit in predicting processed subitltes to be processed.

The output computation from the *glmer* fits through maximum likelihood in R is showed in Figure 6.9 in Appendix D. The exponential of the coefficient for the English subtitles is 2.28, which means the odds for the English subtitles being processed were about 128% higher than the odds of the Chinese subtitles being processed by Chinese participants (ratio of the odds for the English subtitles being processed was 1:2.28), but the difference is not significant ($p > 0.05$).

A line graph was plotted in Figure 6.10 to show the percentage of skipped subtitles among the five videos for CS and ES. The line graph shows that video 4 were most skipped (17.24%) and video 2 is least skipped (10.36%) in the English subtitles, while the skipping rate for the Chinese subtitles (20–25%) is relatively consistent for all videos.

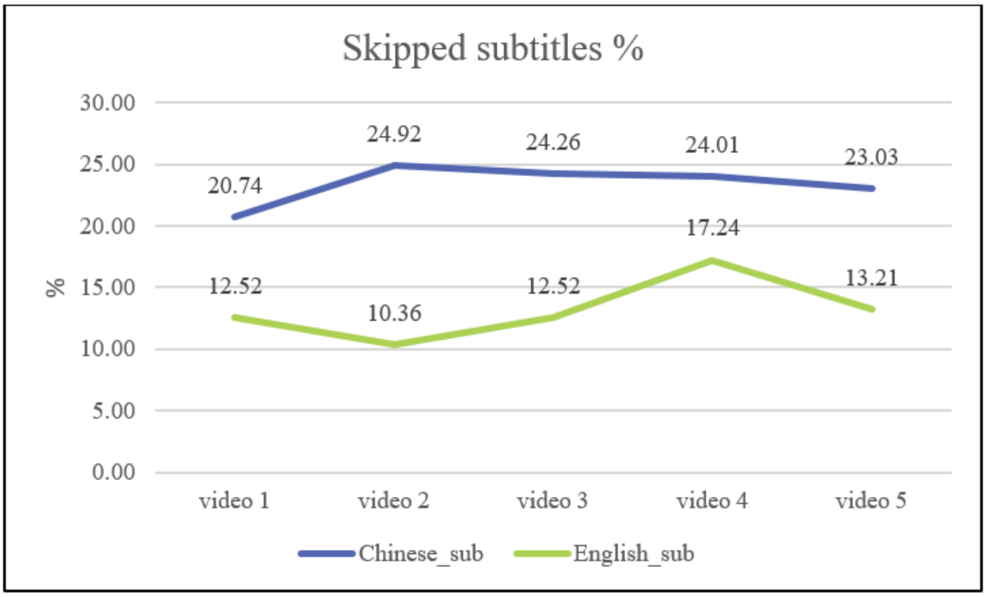


Figure 6-10. Skipped subtitles (%) by Chinese participants.

The content comparison between video 2 and 4 is presented in Table 6.3 (refer to *Figure 3.4* for content comparison between all videos). Video 2 consists of 54% NRN type of redundant (non-redundant with no hand movement) and 27% RH redundant (redundant with hand movement), and video 4 consists of 10% NRN and 40% RH redundant content. Figure 6.10 and Table 6.3 indicate that English subtitles that are redundant with hand distraction were skipped relatively more than English subtitles that are not redundant with no hand distraction.

Table 6-3

Summary table compares the content of video 2 and video 4 in terms of redundant information

Redundant type	Video 2	Video 4
NRN	54%	19%
NRWH	16%	12%
RH	27%	40%
RN	4%	29%

A correlation was calculated between subtitle language, skipped subtitles, redundancy, CL ratings and comprehension scores for all five videos. Results showed that NRN and RN type of redundant content correlate significantly with skipped English subtitles. The output script of the correlation can be found in Figure 6.11 in Appendix D. The correlation showed that skipped English subtitles and NRN redundant content is strongly negatively correlated, $r(942) = -0.96$, $p < 0.05$; and skipped English subtitles and RN redundant content is strongly positively correlated, $r(942) = 0.98$, $p < 0.05$.

Table 6-4

Correlation between redundant type and skipped Chinese and English subtitles for 5 videos

Redundant type	Skipped Chinese subtitles	Skipped English subtitles
NRN	0.27	-0.96 *
NRWH	-0.79	-0.39
RH	-0.07	0.75
RN	0.17	0.98 *

Note: * $p < 0.05$.

6.5.2 Eye tracking measurements

Following the binomial modelling of the eye tracking data in determining the odds ratio of subtitles being processed or skipped, this section presents the results of fixation counts, mean fixation duration, dwell time percentage, time to first fixation and revisits by using *GLMM* to investigate the behaviour of the eye movements. The interpretation of the outcome will be further discussed in *Chapter 7*.

6.5.2.1 Fixation count (FC)

A *GLMM* analysis was carried out after the data has been prepared. Five models were considered, and the output script of the model comparison is showed in Figure 6.12 of Appendix D. The output shows that fit7 is the best fitted model for the data because it is significant with the smallest *AIC* value (45728). Akaike Information Criterion (*AIC*) is an estimator for the relative quality between a selection of models (Akaike, 1974; Burnham & Anderson, 2004). Burnham and Anderson (2004) stated that *AIC* estimates the amount of information lost in a given model, so the model that has lesser information lost is relatively a better fit model. In other words, the model with the smallest *AIC* value is relatively a better model in predicting outcome. The output description for fit7 is presented in the following paragraph.

In this model, *CPS_centred* was added as an interaction to *Group*, and *MW_centred* and *Redundant* as fixed effects. Participant and AOI were added as random effects, they were presented as (1 | Participant) and (1 | AOI). The output script is showed in Figure 6.13 in Appendix D. The results indicated that the difference in fixation counts between ES ($M = 5.19$, $SD = 3.66$) and CS ($M = 3.54$, $SD = 3.34$) is statistically significant ($p < 0.05$, $d = 0.47$); *CPS* and *MW* has a significant impact on fixation counts both with $p < 0.05$; NRWH and RH type of redundant content also has a significant impact on the number of fixations ($p < 0.05$). The results indicated that participants have significantly more fixations in reading English subtitles than Chinese subtitles, which is presented through a boxplot in Figure 6.14. It also showed that when there is hand movement on the screen regardless of redundant information (i.e. redundant content and non-redundant content both with hand movements), fixation counts for reading Chinese subtitles is significantly lower than when there is non-redundant information on screen with no hand movements in the same language subtitles. By contrast, there is no significant difference in fixation counts between redundant and non-redundant information in the absence of hand movements on screen for Chinese subtitles. Furthermore, there is also no significant difference in the number of fixations in the interaction between language and *CPS* after *CPS* has been scaled and centred in both languages.

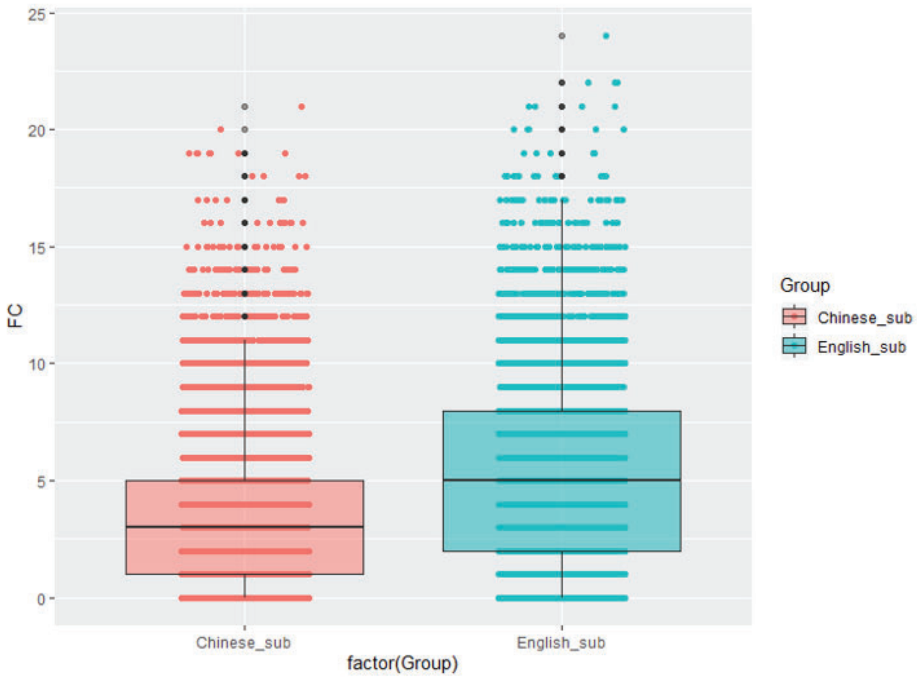


Figure 6-14. Number of fixations for CS and ES.

DHARMa (Hartig, 2019) in R package was used in testing the homogeneity and normality of the residual of this model through simulation. The function *testResiduals* was then used to test uniformity, dispersion and outliers of the residuals of the model.

A QQ-plot and two histograms were generated and showed in Figure 6.15 to provide a visualisation of the outcome of the residual tests. The histogram in the middle of Figure 6.15 shows that the fitted model (red line) lies within the residual simulation, and the histogram on the right shows that the simulated residuals create very little outliers (red cross). Both histograms indicate that the model is a good fit. The QQ-plot on the left of Figure 6.15 shows that the sample is not normally distributed by using Kolmogorov-Smirmov normality test ($p < 0.05$). Noted that the distribution of the sample usually is not expected to be normally distributed due to the nature of eye tracking data.

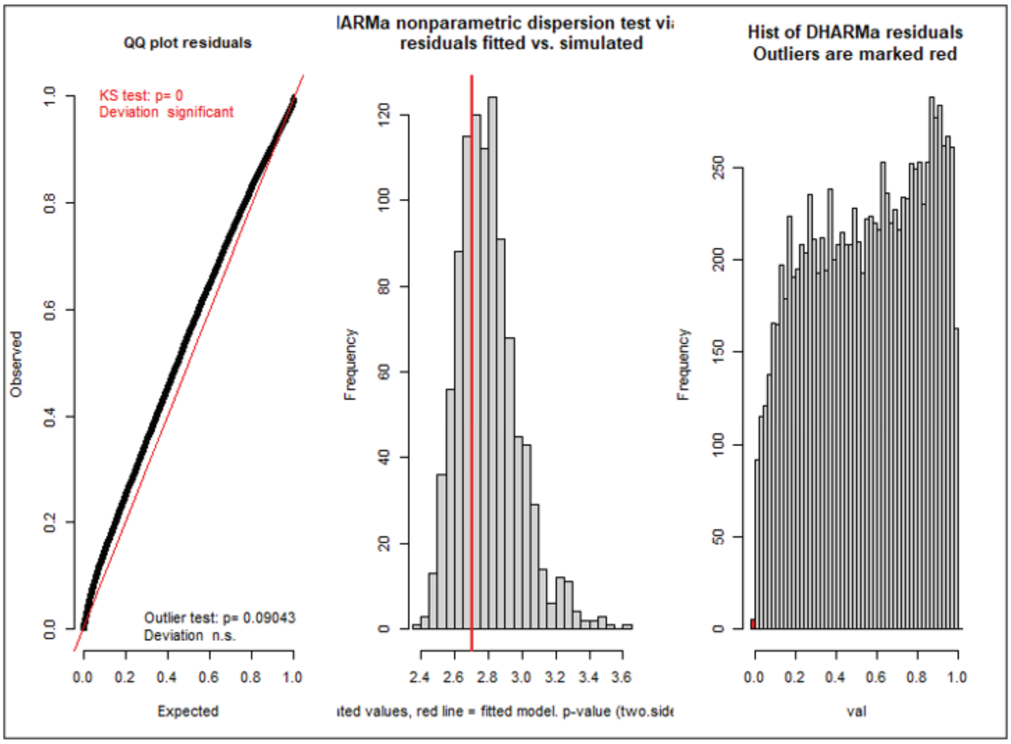


Figure 6-15. DHARMa non-parametric dispersion test in R.

6.5.2.2 Mean fixation duration (MFD)

A *GLMM* was used for MFD analysis. Seven models were considered and are showed in Figure 6.16 in Appendix D. The output script shows that mod6 is the best fit for the data because the model is significant with the smallest *AIC* value (119081). In this model, *Redundant* was factored as a fixed effect and *MW* as an interaction with *Group*. Participant and AOI were added as random effects. The detail output script is showed in Figure 6.17 of Appendix D.

The results showed that the difference in MFD between ES (M = 275, SD = 4.08) and CS (M = 214, SD = 3.62) group is statistically significant ($p < 0.05$, d = 15.82); the difference in MFD in reading the Chinese subtitles when information is non-redundant with hand movement (NRWH) is also statistically significant ($p < 0.05$); furthermore, the average mean word (*MW_centered*) length in the Chinese subtitles has a significant impact on MFD in CS group, and the difference in MFD in the interaction between ES group and the average mean word length in English (*MW_centered*) is statistically significant ($p < 0.05$). The values of the intercepts have been calculated in a detail computation which can be seen in Figure 6.18 in Appendix D.

The outcome of the computation shows that participants have significantly longer mean fixation duration in reading English subtitles (305ms) than Chinese subtitles (266ms). The results also indicated that participants have significantly shorter mean fixation duration in reading Chinese subtitles when the average mean word length in Chinese subtitles is shorter, and when the information is not redundant with hand distraction on the screen. Both yielded the same mean fixation duration (253ms) through the detail computation. Furthermore, the difference in mean fixation duration in the interaction between English subtitles and the average mean word length in English is statistically significant. In other words, participants have significantly longer mean fixation duration (304ms) in reading English subtitles with an increased mean word length than reading Chinese subtitles when the average mean word length in Chinese subtitles is reduced (253ms). The residual test showed uniformity and homogeneity, and there is no observable pattern. The residual graphs are showed in Figure 6.19 - 6.21, and their codes can be found in Figure 6.22 in Appendix D.

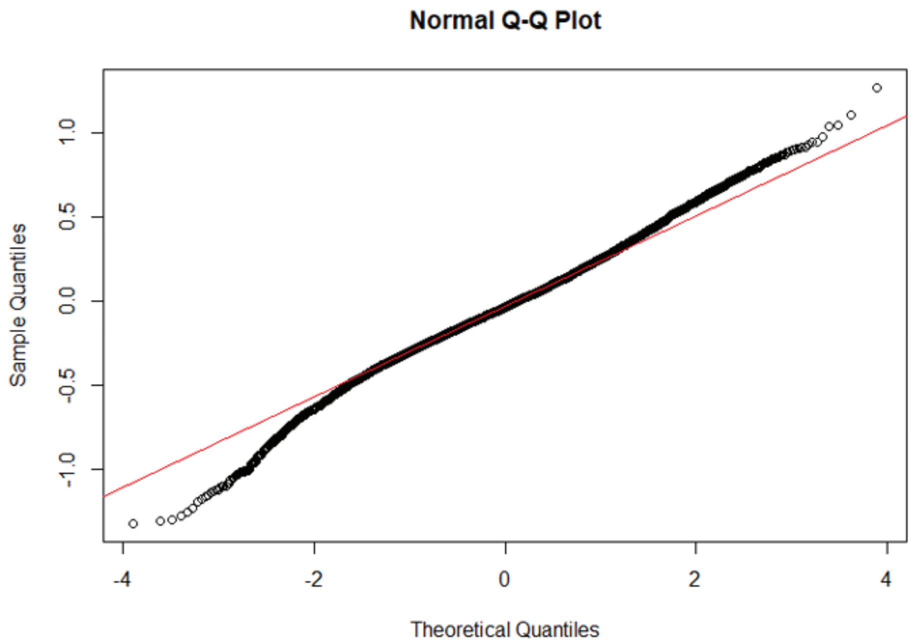


Figure 6-19. QQ-plot showing residual from MFD analysis.

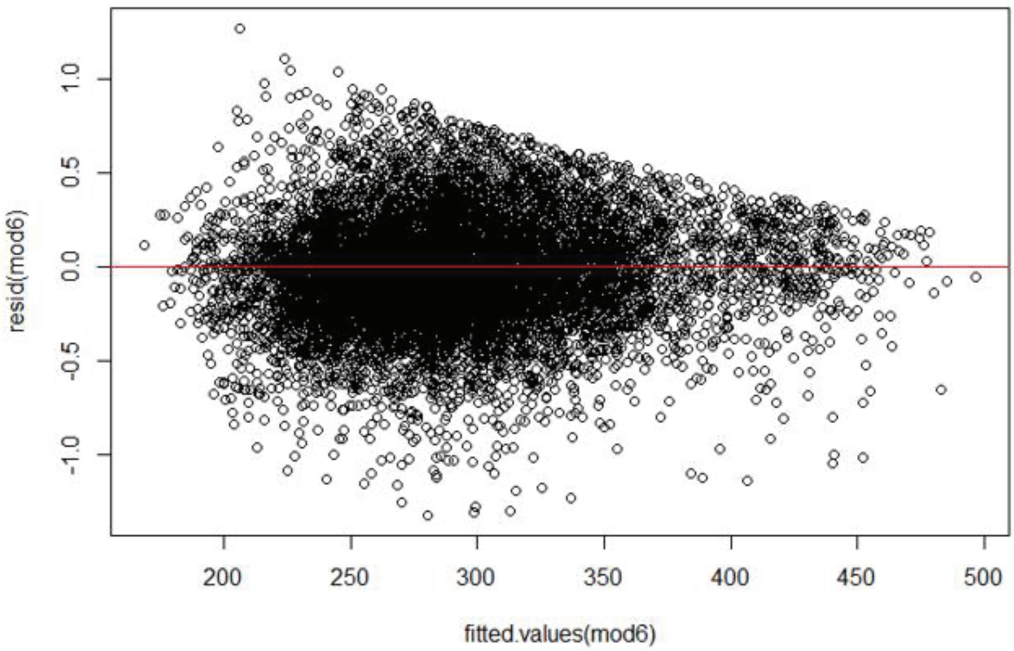


Figure 6-20. Scattered plot showing residual from MFD analysis.

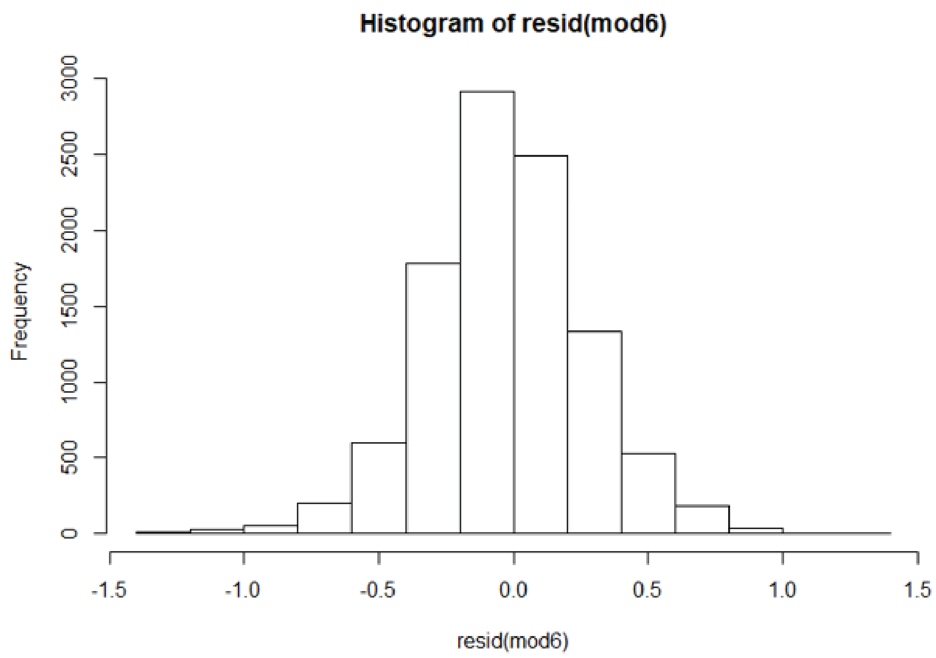


Figure 6-21. Histogram showing residual from MFD analysis.

A boxplot was used to present the average mean fixation duration of the participants in reading English subtitles and Chinese subtitles (Figure 6.23). It can be seen that participants have longer mean fixation duration in reading English subtitles (304ms) than Chinese subtitles (266ms), and the difference is statistically significant.

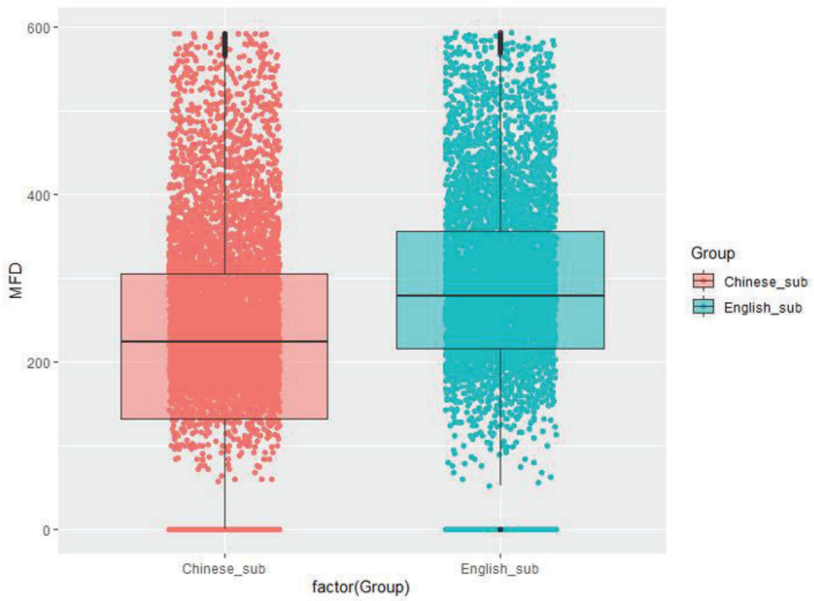


Figure 6-23. Average MFD between Chinese and English subtitles.

6.5.2.3 *Dwell time percentage (DT%)*

Generalised Linear Mixed Models using Template Model Builder (*glmmTMB*) (Brooks et al., 2017) was used for this analysis as an earlier attempt in *glmer* was unsuccessful due to the failure in the residual test. Twelve models were considered and are showed in Figure 6.24 in Appendix D.

The output script shows that fit11 is the best fit for the data because the model is significant with the smallest *AIC* value (88238). In this model, *CPS_centred*, *MW_centred* and *Redundant* were factored as fixed effects. Participant and AOI were added as random effects. The detail output script is showed in Figure 6.25 of Appendix D.

The result showed that the difference in DT% between ES (M = 42.7, SD = 3.95) and CS (M = 27, SD = 3.53) group is statistically significant ($p < 0.05$, $d = 4.19$); *MW* and *CPS* have significant impact on DT% both with $p < 0.05$; NRWH and RH type of redundant content have significant impact on DT% ($p < 0.05$) as well whereas the impact of RN type of redundant content is not significant ($p > 0.05$). The result is presented in the boxplot in Figure 6.26 which shows that participants have significantly longer dwell time percentage in reading English subtitles than Chinese subtitles.

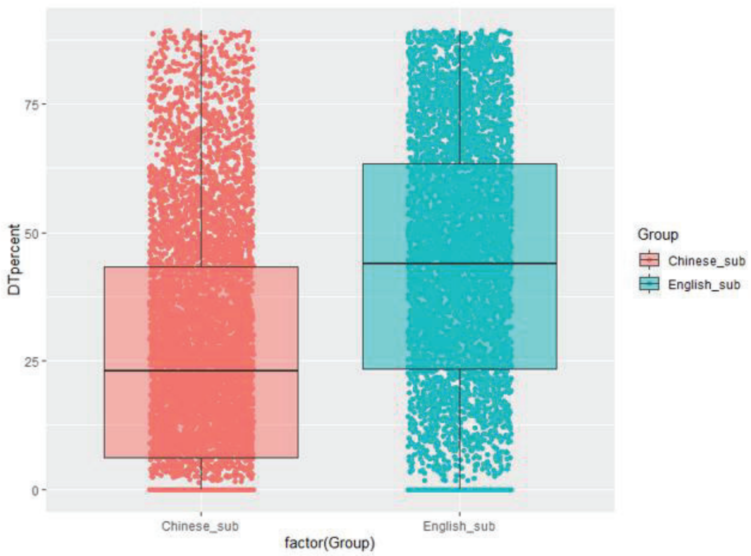


Figure 6-26. DT% between CS and ES.

The result also showed that the presence of hand movement significantly impacted the time participants spent in reading Chinese subtitles regardless of redundant information. Furthermore, the increase of subtitling speed (*CPS*) and Chinese subtitle mean word length (*MW_centered*) have a statistically significant impact on participants in reading the Chinese subtitles. The residual test shows uniformity and homogeneity, and there is no observable pattern. The residual graphs are showed in Figure 6.27- 6.29.

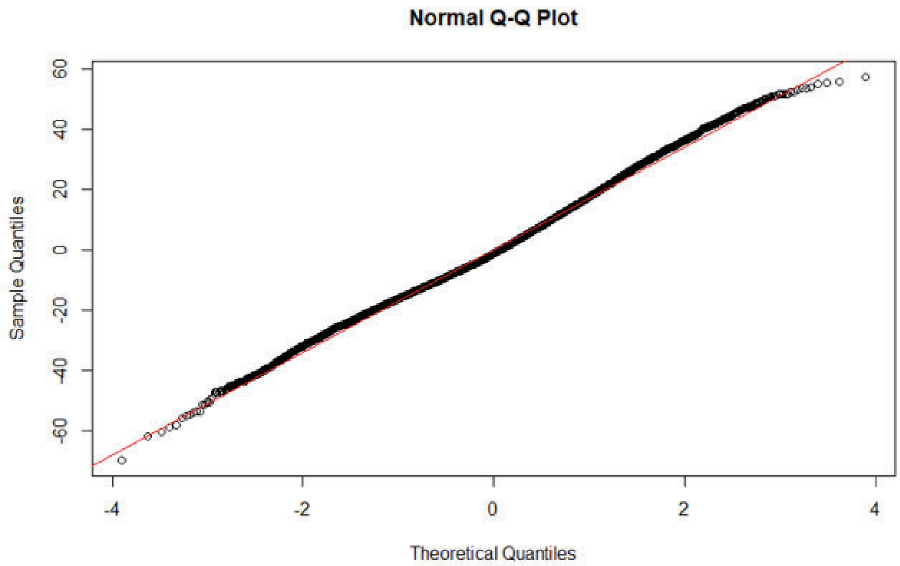


Figure 6-27. QQ-plot showing residual from DT% analysis.

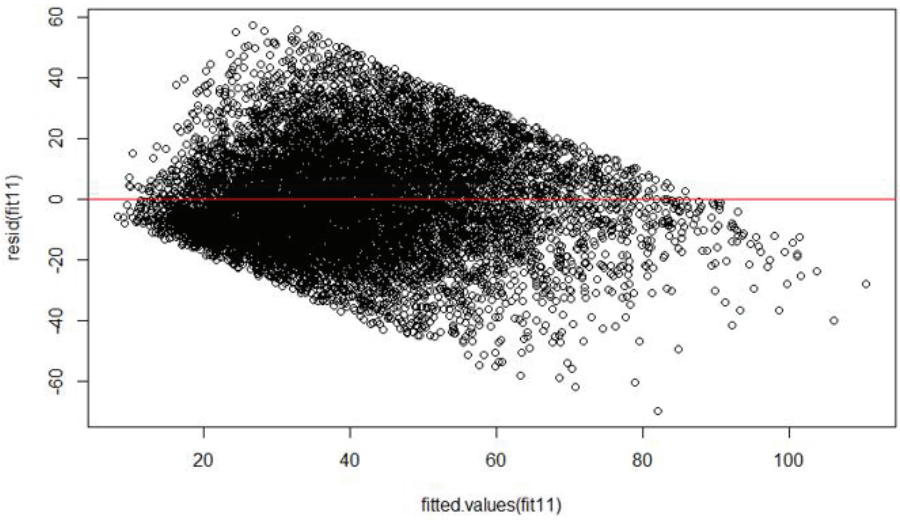


Figure 6-28. Scattered plot showing residual from DT% analysis.

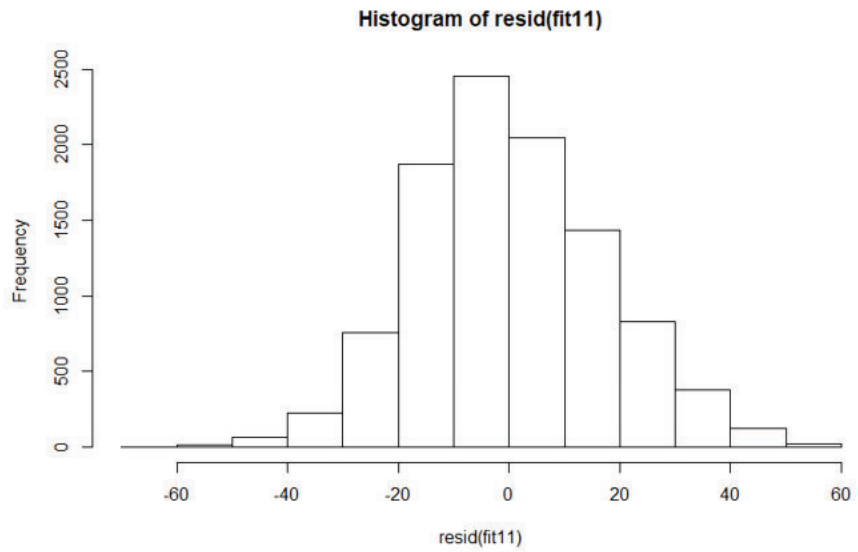


Figure 6-29. Histogram showing residual from DT% analysis.

Stacked bars were plotted to compare the proportion of DT% between ES (Figure 6.30) and CS (Figure 6.31) in terms of redundant information. It can be seen from Figure 6.30 that participants viewing English subtitles spent more time in reading subtitles with no hand movement (NRN and RN) regardless of the redundant information. In other words, English subtitles were read more without the distraction of the hand movement. Dwell time percentage was much lower in CS for all the videos in general. So far, the analysis is indicating that participants have significantly different reading patterns in reading English subtitles than Chinese subtitles with more fixations and longer mean fixation duration, but comprehension of the participants viewing Chinese subtitles have significantly higher scores than those viewing English subtitles.

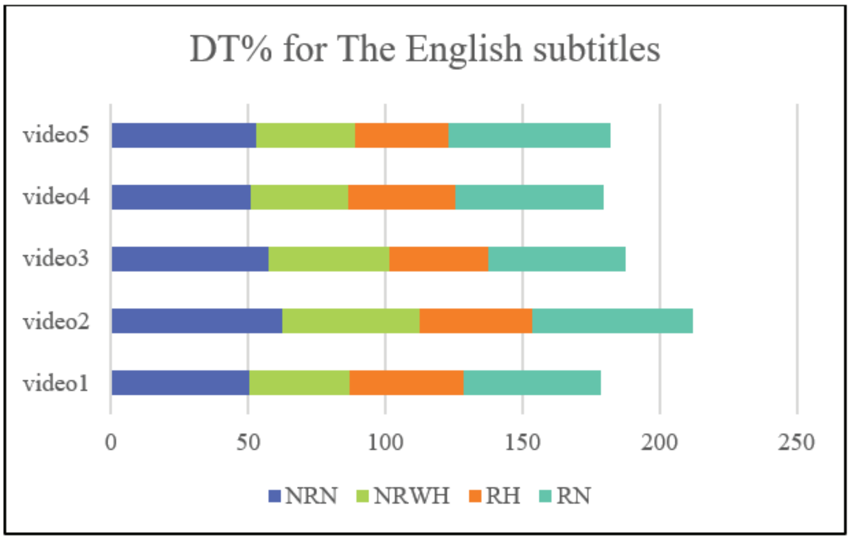


Figure 6-30. Proportion of DT% in ES for redundant information.

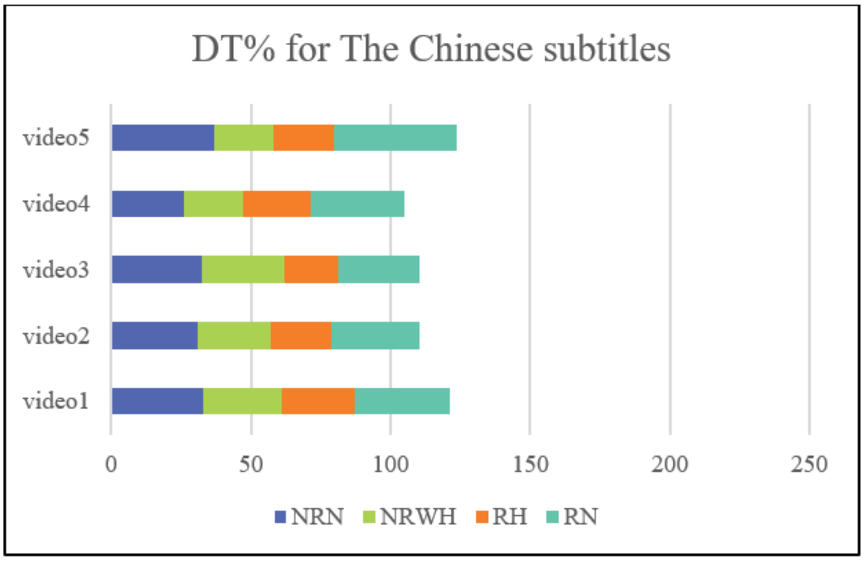


Figure 6-31. Proportion of DT% in CS for redundant information.

6.5.2.4 Time to first fixation (TTFF)

GLMM was used to analyse TTFF. The data was cleaned and centered before performing a *glmer* analysis (see Figure 5.3). Three models have been considered and their comparison is presented in the output script in Figure 6.32 of Appendix D.

The results showed that model fit2 is the best fit with the smallest *AIC* (157823) and $p < 0.05$. In this model, *Redundant* was factored as a fixed effect to *Group*, with Participant and video/AOI as random effects. The results from the output script in Fig 6.33 in Appendix D shows that the difference in TTFF for NRWH (M = 639, SD = 3.35) and RH (M = 720, SD = 3.74) type of redundant in CS are statistically significant (t (5047) = 10.6, $p < 0.05$, d = 22.81); but there is no significant difference ($p > 0.05$) between ES and CS, and in RN type of redundant content. In other words, subtitle language has no significant impact in affecting the duration to the first fixation on the AOI, but the presence of hand movement is a major factor affecting TTFF regardless of the redundant information.

6.5.2.5 Revisits

Two graphs were plotted from the ES and CS group for *CPS* and *CPS_centred* at 99% quantile to examine the number of revisits. These two graphs (Figure 6.34 and Figure 6.35) clearly show the relationship between subtitling speed (*CPS*) and revisits, and the importance to normalise between the two languages.

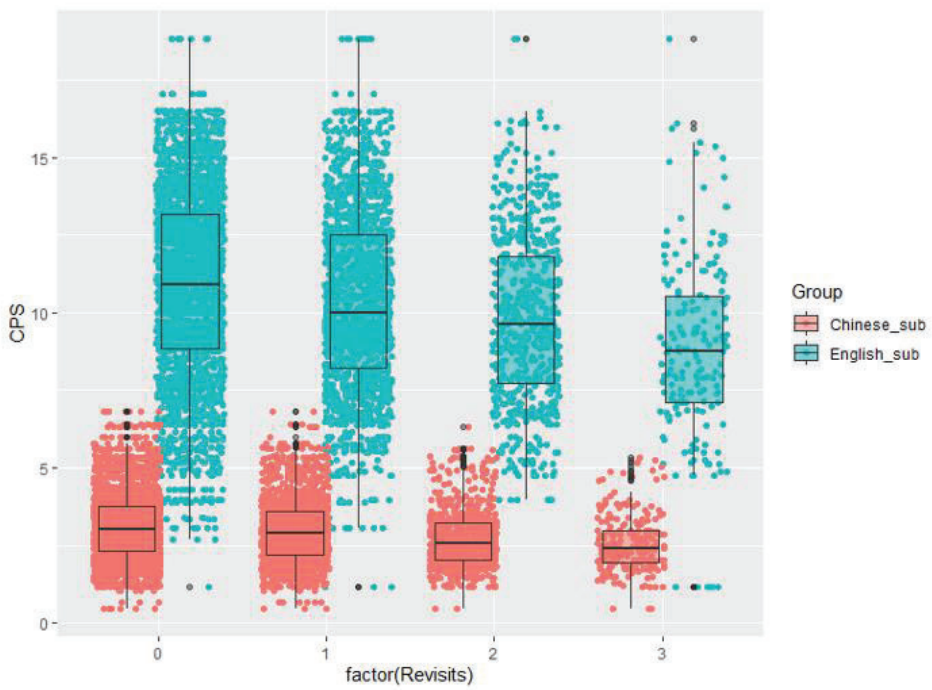


Figure 6-34. Revisits plot for CS and ES before *CPS* was scaled and centred.

First, the graphs show that less revisits occur with an increase of subtitling speed due to the fact that there is not enough time to re-read the subtitle as the speed increases. Second, CS has smaller *CPS* than ES, which is expected as English has longer average word length than Chinese, so the subtitling speed for Chinese is always slower than those in English. However, in Figure 6.35, the *CPS* for both languages are almost the same after *CPS* has been normalised by scaling and centring.

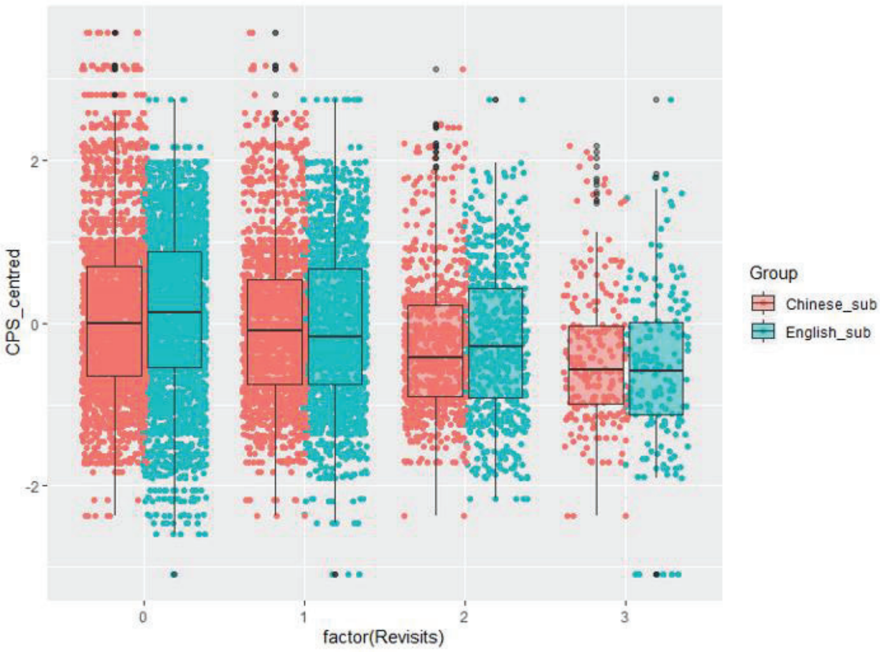


Figure 6-35. Revisits plot for CS and ES after *CPS* has been scaled and centred.

No further computation was needed as the graphs have already provided the necessary answer to the relationship between subtitling speed and languages with an assumption that redundant information plays little part in revisits.

6.5.3 Relationship between comprehension and eye tracking measures

A Pearson correlation was performed using IBM SPSS Statistics 26 to analyse the relationship between comprehension scores and eye tracking measures including mean fixation duration, number of fixations, and dwell time percentage for Chinese subtitles group and English subtitles group. The results showed that the correlations are not statistically significant.

6.5.4 Subtitle processing plots for ES and CS

The following plots (Figure 6.36 – Figure 6.40) provide a visualisation of the way English (green) and Chinese (blue) subtitles were being processed. It can be seen from these five plots that participants generally processed English subtitles and skipped Chinese subtitles, which has been showed in the earlier section through the binomial modelling with no significant difference (see section 6.5.1).

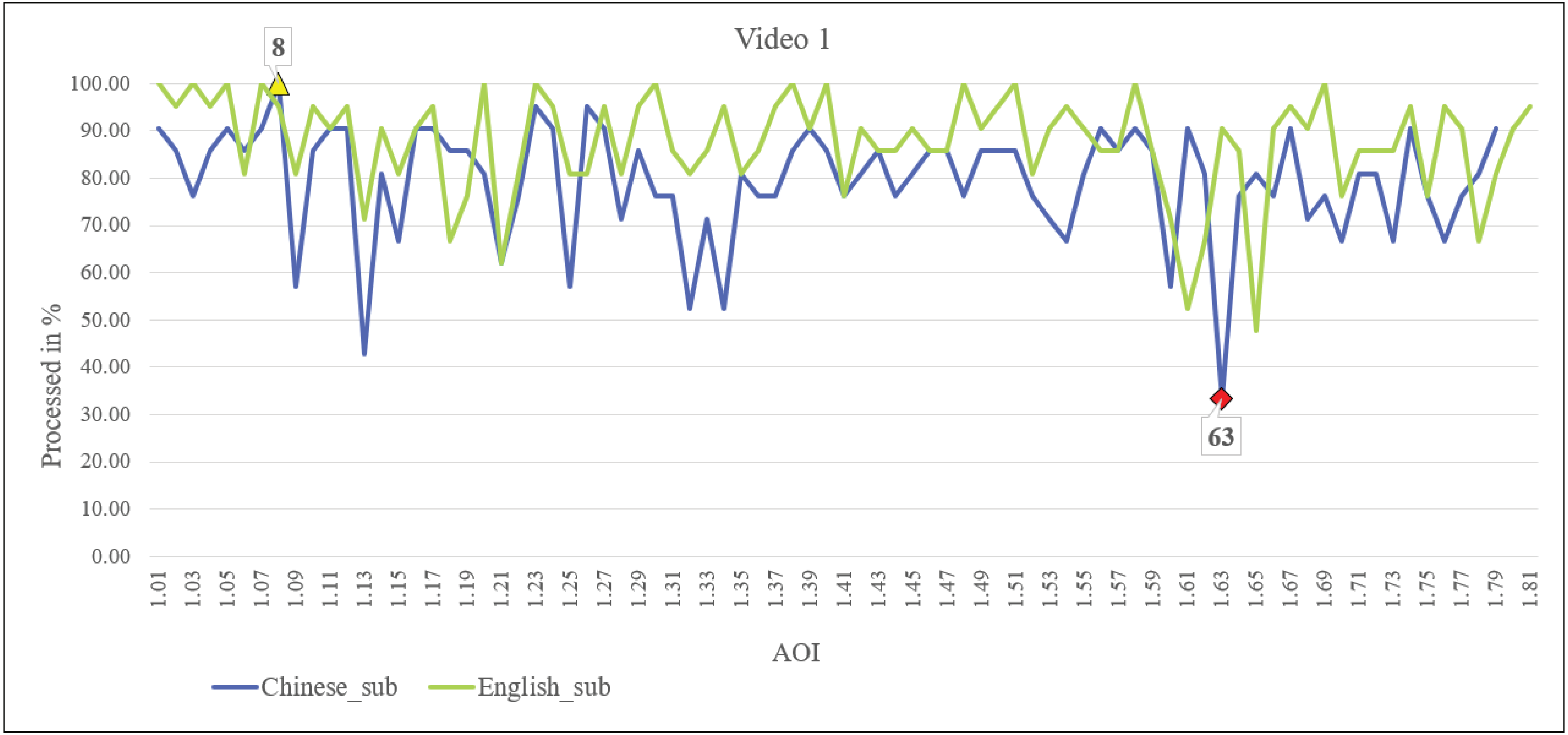


Figure 6-36. English and Chinese subtitles processed plot for video 1.

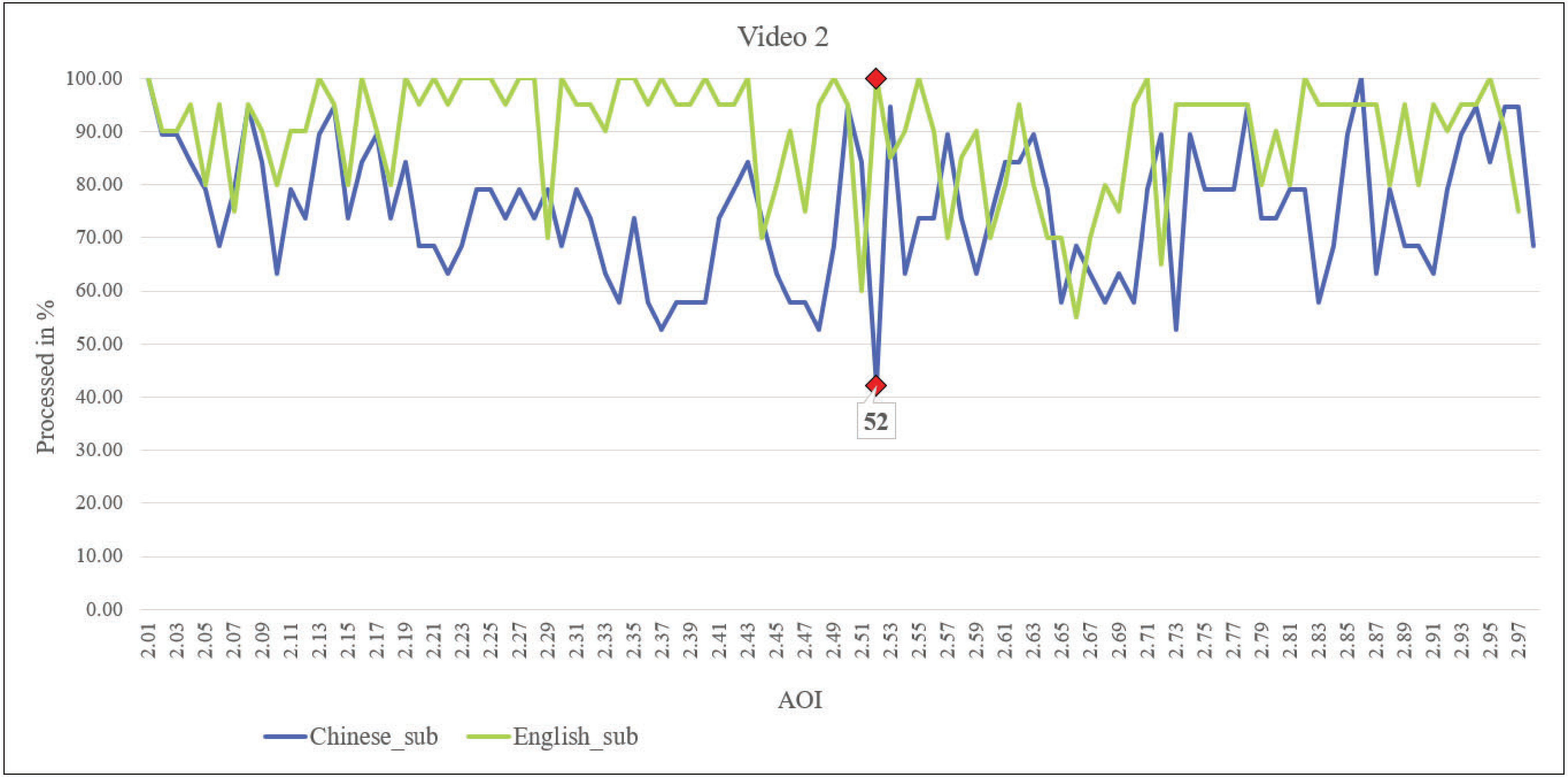


Figure 6-37. English and Chinese subtitles processed plot for video 2.

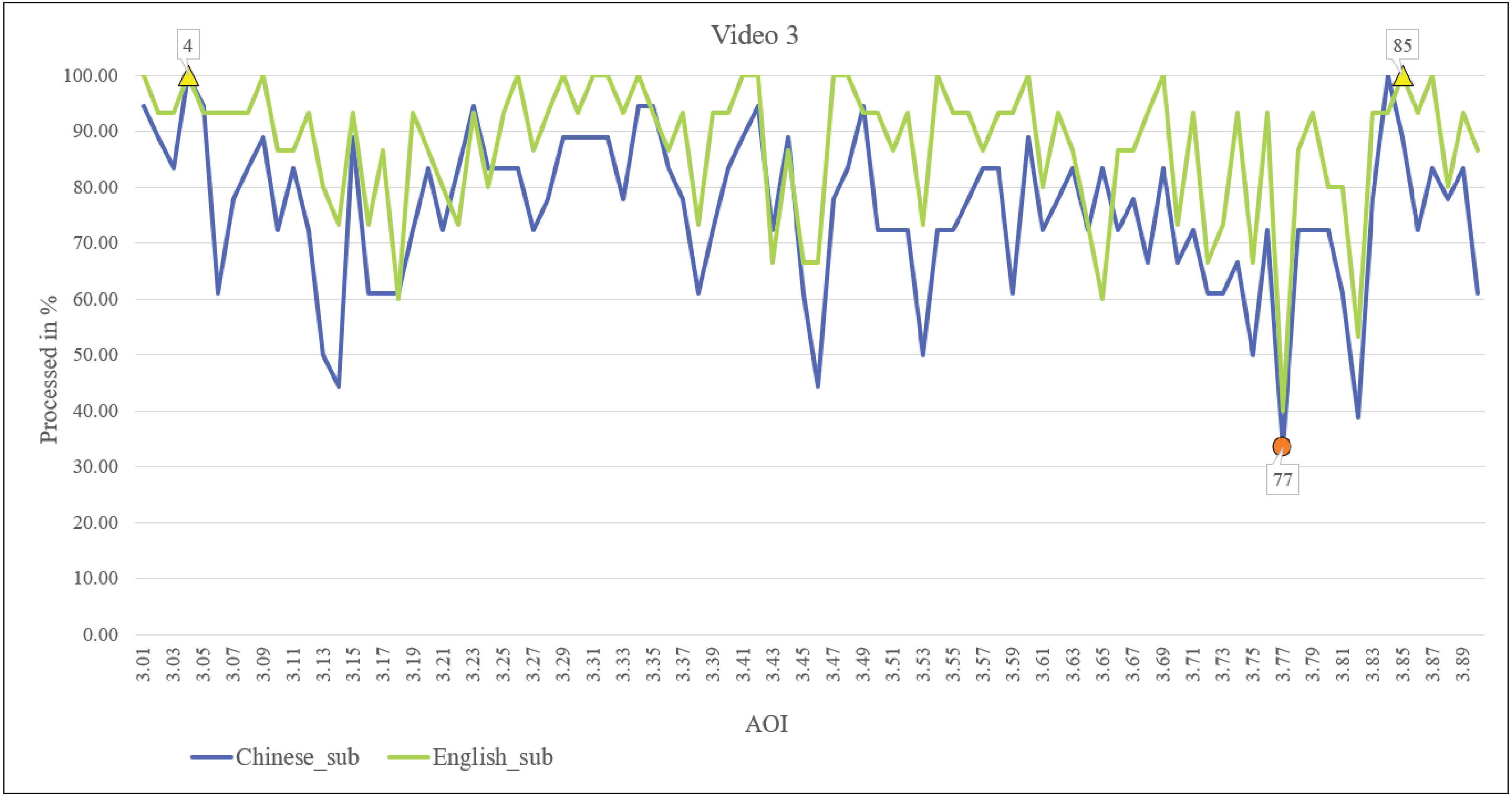


Figure 6-38. English and Chinese subtitles processed plot for video 3.

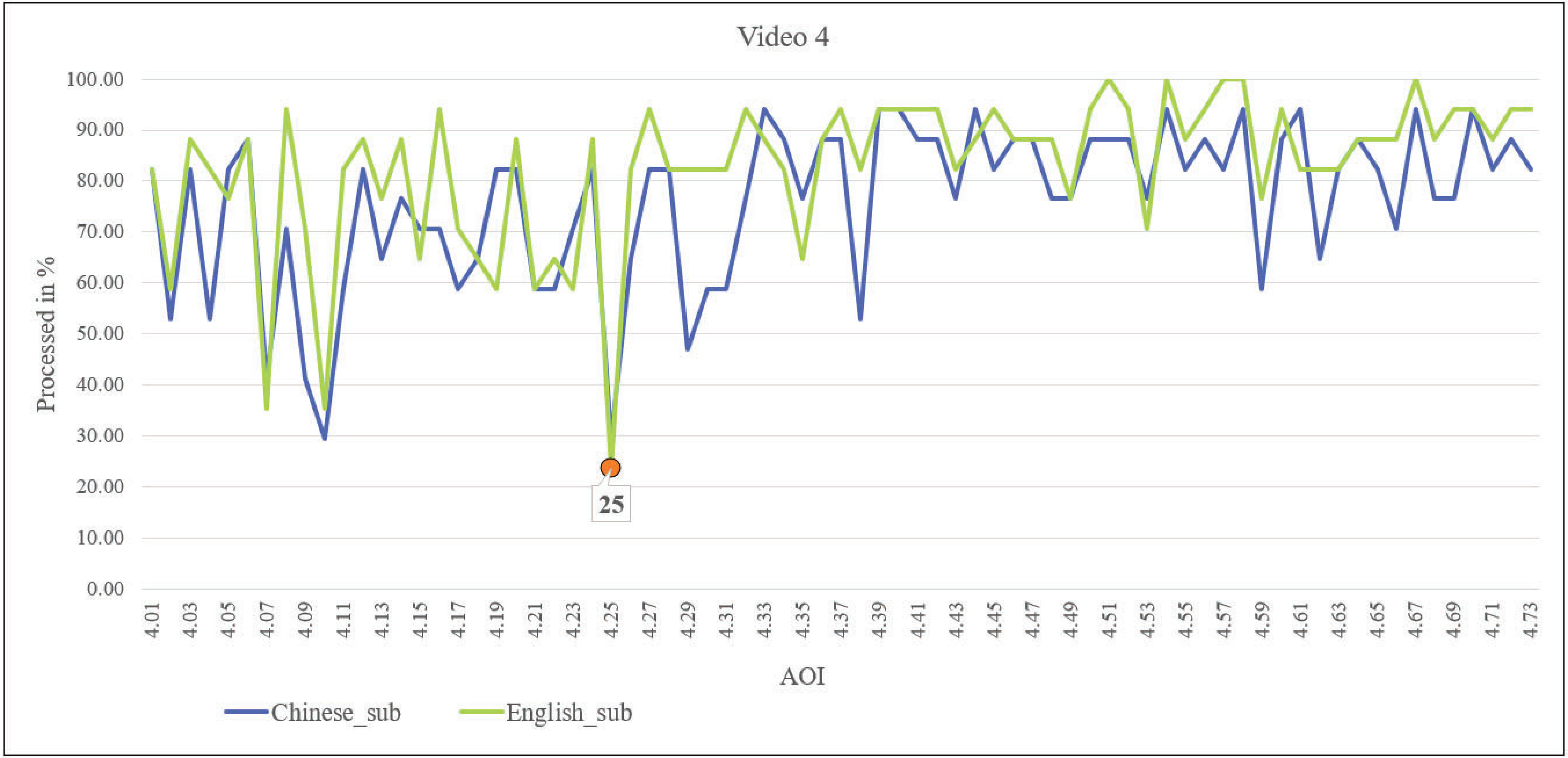


Figure 6-39. English and Chinese subtitles processed plot for video 4.

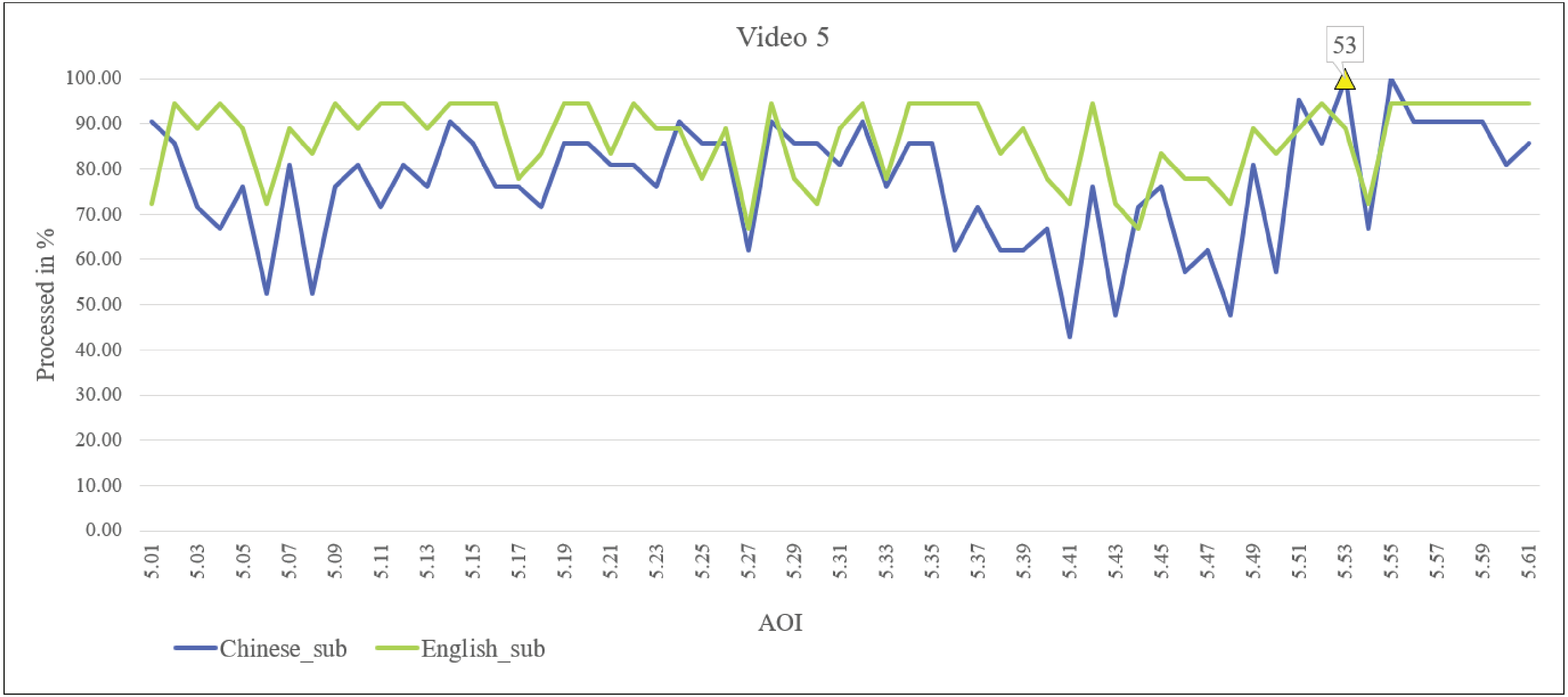


Figure 6-40. English and Chinese subtitles processed plot for video 5.

6.5.5 Heatmaps for AOIs

This section presents a few heatmap comparisons from a number of chosen AOIs to provide a visualisation of how participants distributed their attention on the screen in the presence of English or Chinese subtitles. Heatmaps on the left is an English subtitle with an AOI tag being visible, whereas a Chinese subtitle is showed on the right with no visible AOI tag.

Figure 6.41 and 6.42 show AOI 1.63 and AOI 2.52 where more than half of the participants processed ES and skipped CS (refer to Figure 6.36 and Figure 6.37). It can be seen from both AOIs that the English subtitle is longer (2 lines) than that of the Chinese (1 line), which could be the reason for higher processing rate.

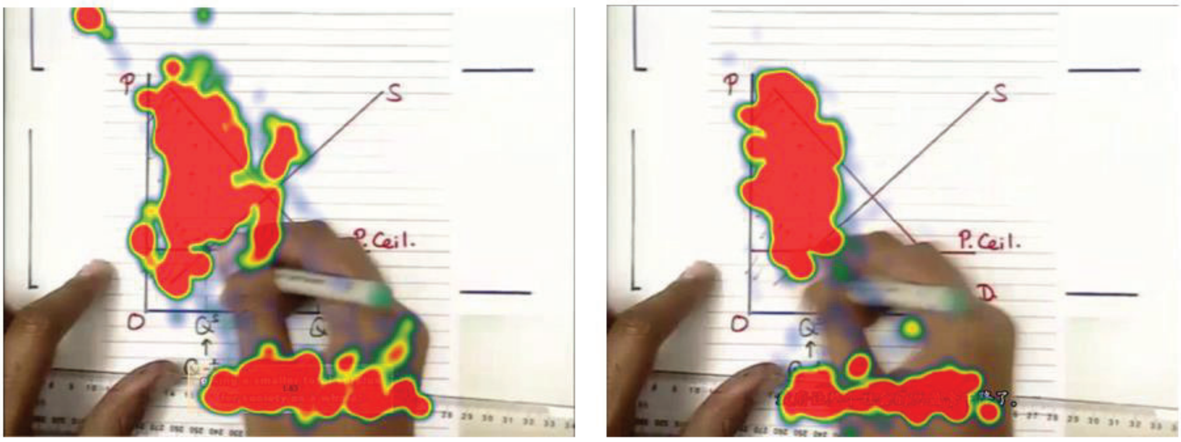


Figure 6-41. ES (left) is processed 57% more than CS (right) in AOI 1.63 with RH redundant.

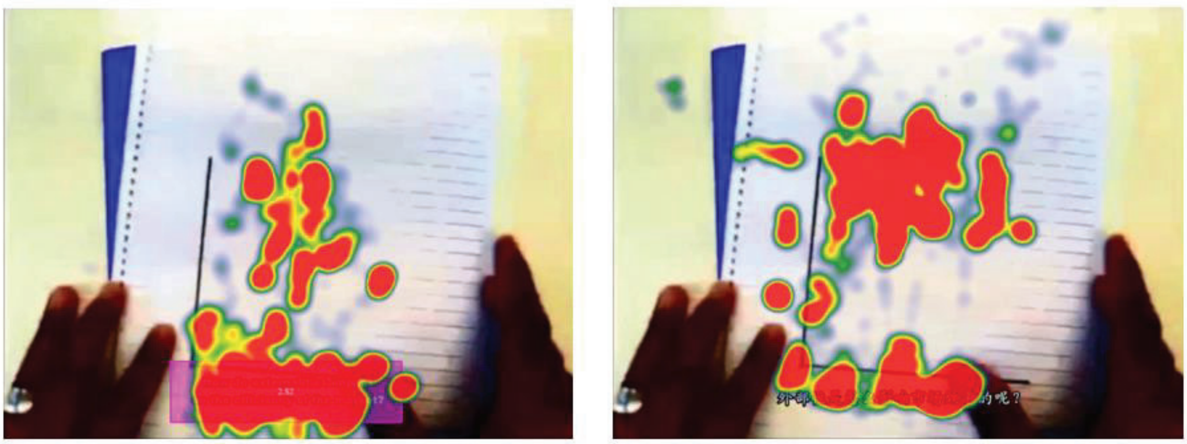


Figure 6-42. ES (left) is processed 58% more than CS (right) in AOI 2.52 with NRWH redundant.

The following heatmaps in Figure 6.43 and 6.44 show low processing rate (33% and 24%) in both CS and ES (refer to Figure 6.38 and Figure 6.39). It can be seen from both heatmaps for AOI 3.77 and AOI 4.25 that low processing rate may be a result of high level of visual complexity, the presence of redundant information and distraction from the hand movement with a relatively short AOI.

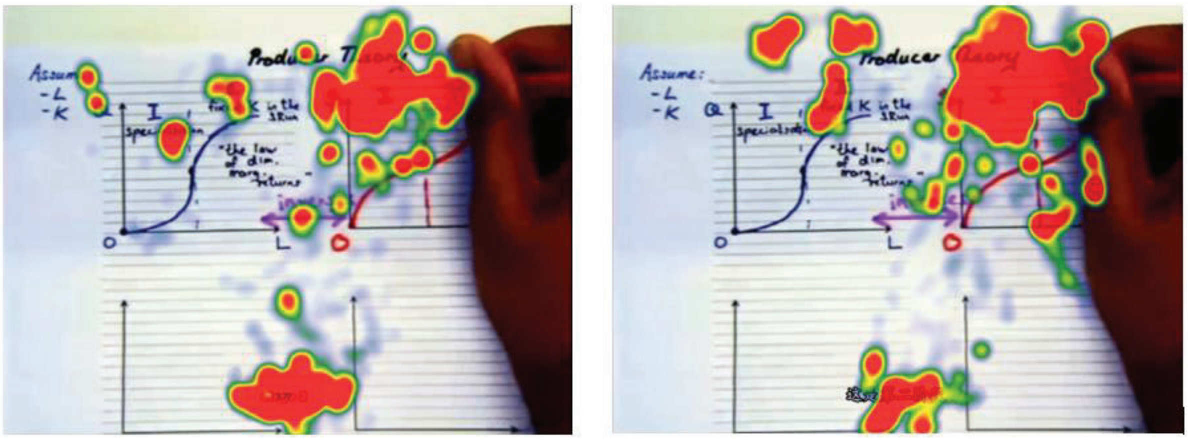


Figure 6-43. 33% of the participants in CS and ES processed AOI 3.77 with RH redundant.

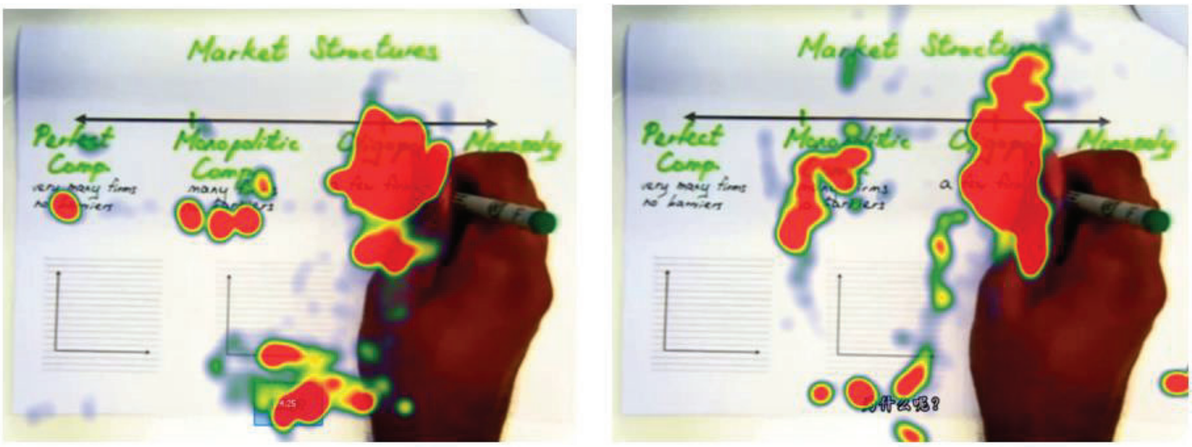


Figure 6-44. 24% of the participants in CS and ES processed AOI 4.25 with RH redundant.

The following heatmaps show that both CS and ES has close to 100% processing rate, that means these AOIs were read by almost all participants (refer to Figure 6.38 and Figure 6.40). There are three possible circumstances that lead to 100% processing rate. First, when the visual complexity on the screen is low, as illustrates in Figure 6.45. The visual image where AOI 3.04

is displayed consists of very little information, so the participants were able to read the subtitle without distraction.

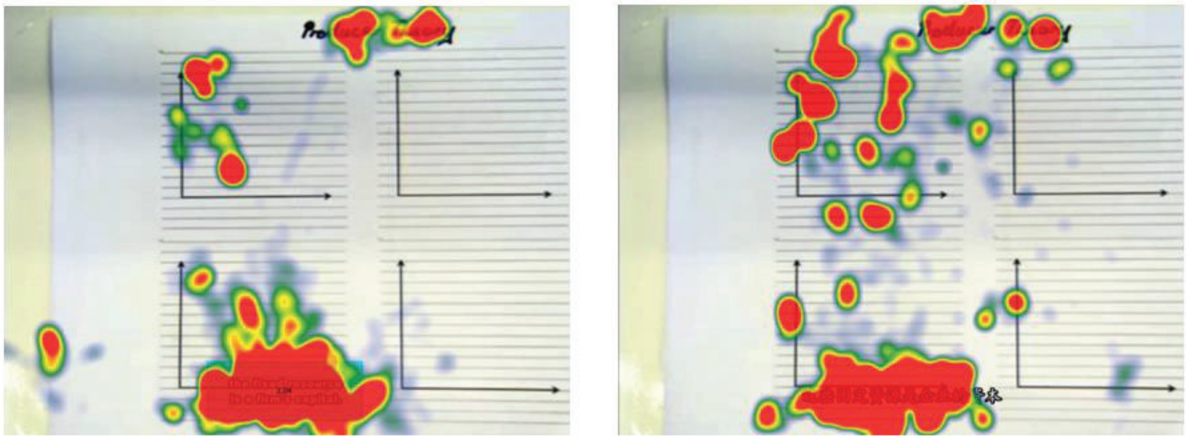


Figure 6-45. AOI 3.04 has a visible time of 4 seconds.

Second, when the visible time of the AOI is long, as shows in Figure 6.46. AOI 5.53 has a visible time of 8 seconds, even though the image on the screen is visually complex, the display time of the subtitle is long enough for the participants to process.

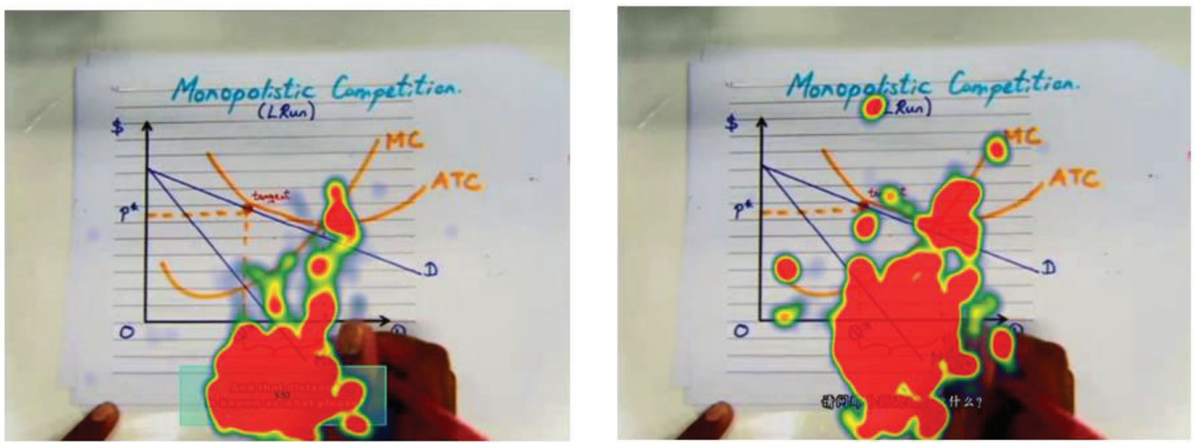


Figure 6-46. AOI 5.53 has a visible time of 8 seconds.

Third, when the same image on the screen has been displayed long enough before the specific AOI was visible, as illustrates in Figure 6.47. AOI 3.85 is a long subtitle with a relatively short visible time of 3 seconds, and the image on the screen is quite complex as well, however, since the image has been visible with emerging on-screen information to the participant from AOI

3.40 onwards, the participants were able to process the subtitle despite the short display time of the AOI.

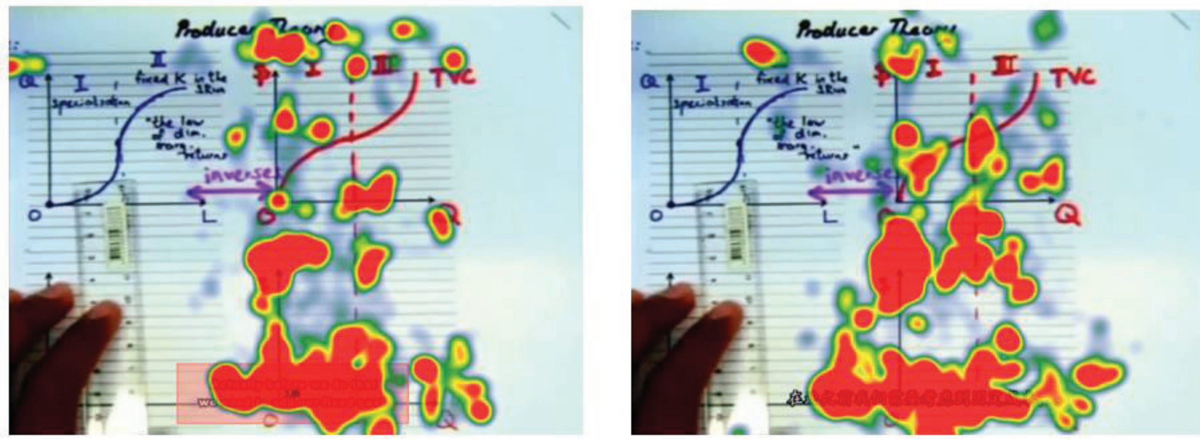


Figure 6-47. AOI 3.85 has a visible time of 3 seconds.

6.6 Summary of results

The results from the analysis in Experiment 2 showed that the level of visual complexity including hand movement on the screen is a major factor affecting the way participants read subtitles, whereas redundant information, in this case, does not seem to impose much impact on audiovisual text processing as it has been proven in previous research (e.g., Homer et al., 2008; Kruger et al., 2013; Liao et al., 2020; Ross & Knowler, 2013). English subtitles with English audio were evidenced to be processed more than Chinese subtitles with English audio in terms of more fixations, longer mean fixation duration, and higher dwell time percentage. However, participants reading first-language subtitles scored higher in comprehension than those viewing second-language subtitles.

Chapter 7 Discussion

7.1 Introduction

Previous research on subtitling mainly focuses on the impact of subtitles and subtitle language on language acquisition and people with hearing loss. There is very limited research on the impact of first- and second-language subtitles, in this case Chinese and English subtitles, on content learning in an educational context. The motivation of the current study is to fill the research gap in investigating the impact of subtitles and subtitle language in audiovisual text processing in content learning for second-language learners in university education. The design of the current study is a mixed-methods approach that is both quantitative and qualitative with objective and subjective measurements. In order to form a robust study, it also incorporates data collected in real-life classroom and laboratory settings. Most of the findings involving Chinese speakers support the hypothesis, but there are some interesting findings involving non-Chinese speakers from the classroom setting that are quite unexpected. The detail interpretation of the current findings is discussed in the following sections, with section 7.2 addressing the impact of subtitles and subtitle language on cognitive load; section 7.3 addressing the impact of subtitles and subtitles language on performance; section 7.4 addressing the impact of subtitle language on audiovisual text processing; section 7.5 addressing the impact of redundant information on audiovisual text processing; section 7.6 discusses the qualitative analysis of the experiment, and section 7.7 provides a summary of the analysis as an overview.

7.2 How do subtitles and subtitle language impact cognitive load?

7.2.1 Experiment 1: Classroom

It is hypothesised that Chinese-speaking participants would have a lower CL in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles, and those reading first-language (Chinese) subtitles would have a lower CL than those reading no subtitles. The results show no significant difference in cognitive load for Chinese speakers reading first-language, second-language or no subtitles, therefore this hypothesis could not be supported. The non-significant results could be explained by various factors affecting the cognitive load for Chinese speakers in reading these three subtitles conditions. When Chinese speakers are reading L1 subtitles, their CL increases because participants are distracted in checking the accuracy of the translation, and at the same time dealing with subject matter with technical language they do not normally use in their studies even if it is their first language. Similarly, L2 subtitles trigger comparison of audio and subtitles; whereas in no subtitles, CL increases as it does not have the visual support that is provided by the other two subtitled conditions.

It is hypothesised that English speaking participants would have a lower CL in reading first-language (English) subtitles than reading no subtitles, and those reading no subtitles would have a lower CL than those reading foreign language (Chinese) subtitles. The results do not support the hypothesis in this case either. The unexpected results show that English speakers have a significantly lower cognitive load when viewing foreign language subtitles than viewing first-language subtitles. It is a medium size effect for this analysis. The outcome suggests that English speakers automatically look at foreign language subtitles but without being able to process it, therefore their overall exposure to meaningful information is reduced, which could explain the lower self-reported cognitive load. When there are no subtitles, they engage with the visuals in an uninterrupted fashion, which means there is more meaningful information to process resulting in a higher cognitive load. When they see first-language subtitles, there are two sources to process visually, which leads to relatively higher cognitive load ratings. Furthermore, English speakers may not have the same level of familiarity as Chinese and foreign language speakers in subtitle reading, therefore the presence of first-language subtitles might create extra cognitive load due to the lack of habitual behaviour and strategies in reading subtitles (see also, Dana, 2004; Vanderplank, 1988). However, this possible explanation will have to be confirmed in a future study.

The results further indicate a significant interaction between subtitle language and the first-language of the participant. The results show that English speakers reading first-language subtitles have a significantly higher cognitive load than Chinese speakers reading second-language subtitles. It is a medium size effect for this analysis. The findings do not support the hypothesis that reading first-language subtitles with first-language audio would create less cognitive load. One of the possible explanations could be that English speakers have not formed a habitual behaviour in subtitle reading as explained earlier, thus even reading first-language subtitles creates relatively more cognitive load. Another possible explanation could be a result of expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003), that the presence of first-language subtitles creates more distraction to English speakers when they are proficient enough to understand the audio information without the subtitles, thus the presence of first-language subtitles actually distracts them in processing the information as they may try to check the accuracy between audio and visual text information.

Even though the research question focuses on English and Chinese speakers, in order to fully utilise the existing data, an additional analysis has also been conducted to investigate how non-Chinese and non-English (foreign language) speakers behave in the current study in terms of cognitive load when compared to Chinese and English speakers. The results show that foreign language speakers have the highest cognitive load when reading foreign language (Chinese) subtitles and lowest cognitive load when reading no subtitles. The results indicate that foreign language participants behaved similarly as the Chinese-speaking participants when they see English subtitles since English is their second-language, so having to process two sources (image and subtitles) visually creates some cognitive load. When there are no subtitles, they engage with the visuals in an uninterrupted fashion, which means they can process more meaningful information, which could explain the significantly low cognitive load. However, the results show some discrepancy in self-report CL between English and foreign language speakers in reading foreign language subtitles, in which English speakers reported the least cognitive load whereas foreign language speakers reported the highest cognitive load. This discrepancy could be explained by the unequal number of participants in Chinese (foreign language) subtitles group in which there are 15 English-speaking participants and only 6 foreign language participants.

However, due to an unequal number of participants in each subtitle condition in Experiment 1, especially between English and foreign language group, the outcome of the analysis involving foreign language speakers in cognitive load lacks statistical power.

7.2.2 Experiment 2: Laboratory

It is hypothesised that Chinese-speaking participants would have a lower CL in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles, and those reading first-language (Chinese) subtitles would have a lower CL than those reading no subtitles. As in Experiment 1, the results show no significant difference in cognitive load for Chinese speakers reading first-language, second-language or no subtitles, therefore this hypothesis could not be supported. The non-significant results could also be explained in the same way for the CL results for Chinese participants in the classroom environment discussed in section 7.2.1.

7.3 How do subtitles and subtitle language impact performance?

7.3.1 Experiment 1: Classroom

It is hypothesised that Chinese-speaking participants would have higher comprehension in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles, and those reading second-language subtitles would have higher comprehension than those reading no subtitles. As mentioned earlier (Chapter 3), the validity of the comprehension scores from the classroom experiment is not a reliable indication of the extent to which the participants benefited from the subtitled videos since the scores were directly taken from the end of term examination, and most of the examination materials were not covered in the videos used in the study. Therefore, the interpretation of the comprehension analysis in the classroom setting can only act as a reference for this study and a guide for further investigation. The comprehension results, under these circumstances, do not support the hypothesis. The results show that the difference in comprehension for Chinese speakers between reading first-language, second-language or no subtitles in the classroom setting is not statistically significant.

It is hypothesised that English-speaking participants would have higher comprehension in reading first-language (English) subtitles than those reading no subtitles, and those reading no subtitles would have a higher comprehension than those reading foreign language (Chinese) subtitles. The results support the hypothesis that English speakers scored significantly higher in reading first-language than foreign language subtitles. The effect size for this analysis is large. For the English-speaking participants who were viewing foreign language subtitles, they still allocate extensive attention to the subtitles in a language they do not understand as a result of automatic reading behaviour (d'Ydewalle & De Bruycker, 2007; d’Ydewalle et al., 1991), which may distract them from comprehending the actual content, and therefore lower comprehension scores. The results indicate that reading first-language subtitles assists comprehension and thus improves performance.

7.3.2 Experiment 2: Laboratory

It is hypothesised that Chinese-speaking participants would have higher comprehension in reading second-language (English) subtitles than those reading first-language (Chinese) subtitles, and those reading second-language subtitles would have higher comprehension than those reading no subtitles. The results do not support the hypothesis. The results show that Chinese speakers scored significantly higher in comprehension when reading first-language subtitles than second-language and no subtitles. The effect size in this analysis is small. The results seem to suggest that reading first-language subtitles possesses an advantage over reading second-language subtitles, in which first-language subtitles may provide deeper cognitive processing in comprehending the text. In other words, Chinese participants seem to comprehend more effectively in reading first-language subtitles, which could also have facilitated a more efficient processing of the visual information in the videos. These results agree with the findings by Kruger et al. (2014). In their study Sesotho students allocated more attention in reading English (second-language) subtitles but have a higher retention of knowledge when reading Sesotho (first-language) subtitles. Kruger et al. (2014) suggested that reading first-language subtitles may provide cognitive priming and therefore benefit first-language speakers in deep cognitive processing.

7.4 How does subtitle language impact audiovisual text processing?

It is hypothesised that participants would allocate relatively less attention in reading first-language subtitles (Chinese subtitles) than in reading second-language subtitles (English subtitles). The binomial modelling does not support the hypothesis as the results show that the attention allocation in reading first- and second-language subtitles is not statistically significant, but the five eye tracking measures including fixation counts, mean fixation duration, dwell time percentage, time to first fixation and revisits support the hypothesis.

The binomial modelling results show that the odds of second-language subtitles being processed is around 128% higher than the odds of first-language subtitles being processed, but the result is not statistically significant. Despite the insignificant odds ratio, the eye movement findings show that Chinese speakers have significantly more fixations counts, longer mean fixation durations and higher dwell time percentage in reading second-language subtitles than first-language subtitles. In other words, with similar processing probability, the eye movement patterns in reading first-language subtitles is significantly different from reading second-language subtitles. Previous reading research showed that information is acquired during fixations (e.g., Holmqvist et al., 2011; Rayner, 1998; Rayner et al., 2006; Rayner, Juhasz et al., 2005; Schotter & Rayner, 2012), and based on the findings of Rayner et al. (2006) that longer fixations, shorter saccades and more regression is an indication of processing complicated information, it is logical to conclude that it is more challenging for Chinese speakers reading second-language subtitles than reading first-language subtitles resulting in more fixations, longer mean fixation duration and longer dwell time percentage. Furthermore, in the presence of hand movements, Chinese speakers took significantly longer time to make their first fixation in the subtitle, regardless of whether the visual content was redundant with the subtitle/speech, and regardless of the language of the subtitle. The effect size for the analysis of mean fixation duration, dwell time percentage, and time to first fixation is large, with a medium effect size for fixation counts.

Integrating eye movement results with the comprehension and the cognitive load analysis, the overall findings suggest that Chinese speakers reading first-language subtitles improved performance; and that their reading patterns change while reading second-language subtitles with more fixations, longer mean fixation duration and longer dwell time percentage. Results from previous studies suggest that second-language audio with second-language subtitles benefits language acquisition most, including listening comprehension, vocabulary learning, and information retention (e.g., Baltova, 1999; Hayati & Mohmedi, 2009; Mitterer & McQueen, 2009). This can be explained by form-meaning mapping (Winke et al., 2010; Winke et al., 2013), meaning that the presence of subtitles assists learners to bridge information between audio and visual presentations, thus improving learning. However, if the purpose of the subtitles is to assist content learning, the ability to understand the concept of the content is more important, and thus the language becomes a tool for learning rather than being the goal of learning. In this case, the results suggest that first-language subtitles may have the advantage of promoting deep cognitive processing of the learning content and lead to better comprehension. The results support the findings by Kruger et al. (2014) that first-language subtitles are beneficial in assisting cognitive processing and thus improving performance.

7.4.1 Comparing reading patterns between Chinese and English text: Static vs dynamic

Previous research found that the average fixation duration of Chinese and English readers in reading first-language text are very similar despite the distinct difference in their language system and visual forms (e.g., Feng et al., 2009; Rayner, 2004; Rayner, Li et al., 2005; Schotter & Rayner, 2012; Sun & Feng, 1999; Sun et al., 1985). Most of the past results seem to agree that the average fixation duration in reading Chinese and English ranges between 225ms and 250ms (Feng et al., 2009; Rayner, 2004; Rayner, Li et al., 2005), with some being more specific with mean fixation duration of reading Chinese as 260ms and English as 270ms respectively (Schotter & Rayner, 2012; Sun & Feng, 1999; Sun et al., 1995). Comparing the current results to that of the previous findings, it is obvious that there is a slightly different outcome in the context of dynamic text, especially on the mean fixation duration of Chinese speakers reading second-language subtitles. The current findings indicate that Chinese speakers have a mean fixation duration of 304ms in reading English subtitles compares to the previous findings of 270ms in reading English static text, while the mean fixation duration of reading Chinese subtitles (266ms) is almost the same when comparing with previous findings of 260ms in reading Chinese static text. The results seem to indicate that reading patterns change when Chinese speakers read second-language subtitles compares to first-language subtitles. It seems to suggest that when second-language text is presented dynamically as subtitles in a multimedia learning environment, it creates extra cognitive load for second-language learners and therefore leading to longer mean fixation duration in order to fully comprehend the content. These results could have important implications for education pedagogy and potential teaching instruction in supporting academic success for second-language learners. Given that the current results provide evidence that L1 subtitles assist in improving comprehension, L1 subtitles could be utilised with great effect in contexts like Australia, the UK and the US where large numbers of students study through the medium of English as a second language, whereas L2 subtitles can be presented in assisting language acquisition, which has been proven to be effective in previous studies (Montero Perez et al., 2013).

7.5 How does redundant information impact audiovisual text processing?

The current findings indicate that regardless of the subtitle language, redundant information has limited impact on subtitle reading in terms of the number of fixations, mean fixation duration, dwell time percentage, time to first fixation and the number of revisits. In addition, the results show that redundant information does not have the negative impact on information processing, in this case subtitles, as has been suggested by past studies (e.g., Diao et al., 2007; Diao & Sweller, 2007; Kalyuga et al., 1999; Kalyuga & Sweller, 2014; Mayer et al., 2001). The results from the eye tracking analysis indicate that the presence of redundant information has no significant impact on the number of fixations, mean fixation duration, dwell time percentage, time to first fixation and the number of revisits. However, the results show that the presence of distractors such as hand movements and the average subtitle word length have more impact on these five measurements regardless of the subtitle language; and the level of visual complexity seems to contribute more factors in processing audiovisual text that has been demonstrated in the qualitative analysis of this study.

Despite the insignificant impact of redundant information on subtitle reading, the results show that there is a high level of relationship between the amount of redundant information and the number of skipped second-language subtitles by Chinese speakers. There are two correlations that are statistically significant with one being positively correlated and the other negatively correlated. The first correlation implies that the more the information is redundant, the more the second-language subtitles are being skipped; the second correlation implies that the more the information is non-redundant, the less the second-language subtitles are being skipped.

7.6 What does the qualitative analysis show?

The eye tracking results in Experiment 2 are different from those obtained in the qualitative analysis in Experiment 1. Experiment 1 suggests that Chinese speakers reading first-language subtitles in an interrupted fashion, but the more robust eye movement patterns in Experiment 2 reveal that Chinese speakers actually read second-language subtitles in an interrupted fashion. This difference could be the result of limited eye tracking sample in Experiment 1 as only four eye tracking datasets are available for analysis. The comparison would not be reliable but for Experiment 1 to indicate that subtitle language has an impact in processing audiovisual text, and to inform a subsequent experiment.

In Experiment 2, the heatmaps provide visualisation of the way Chinese speakers attended to subtitles versus screen. The low processing rate of the subtitles could have resulted from a few factors including high level of visual complexity, the presence of redundant information, distraction from hand movement and a relatively short AOI. The length of the AOI also affects processing rate, 2-line subtitles were observed to have higher processing rate than 1-line subtitles. Future research could include number of lines of subtitles in the investigation. There are three possible circumstances that lead to close to 100% processing rate of the subtitles regardless of subtitle language. First, a low level of visual complexity on the screen. When the visual image consists of very little information, it allows the participants to read the subtitle without distraction. Second, when the visible time of the AOI is long enough for the participants to process regardless of the level of visual complexity. Third, when the same image on the screen has been displayed long before the specific AOI was visible, then the participants are able to process the subtitle despite the short display time of the AOI.

7.7 Summary of the analysis

The results of the current study indicate that the level of visual complexity including the presence of hand movement is a crucial factor affecting the way participants read subtitles in classroom videos where students only see a disembodied hand writing and drawing while the teacher speaks. Redundant information, in this case, has a very limited impact on audiovisual text processing. Chinese speakers processed second-language subtitles significantly differently than first-language subtitles in terms of more fixations, longer mean fixation durations and higher dwell time percentage. Furthermore, Chinese speakers reading first-language subtitles scored significantly higher in comprehension than those reading second-language subtitles. There is high level of correlation between redundant information and second-language subtitles being skipped and the correlation is statistically significant. The effect size of the above analysis is mostly large except for the number of fixations, which has a medium size effect. Furthermore, the current results support previous studies (Schotter & Rayner, 2012; Sun et al., 1985) that reading patterns for static text are similar among different languages because it is the lexical content that is being processed rather than the visual form of the language. However, reading patterns change for reading second-language text in a multimedia learning environment where longer mean fixation duration is needed to properly process the dynamic text. The results resonate with those from Specker (2015) that reading patterns change for second-language speakers in reading second-language subtitles.

There are some interesting findings from the classroom setting involving non-Chinese speakers. One of these findings indicate that English speakers have significantly higher cognitive load in reading first-language subtitles than the foreign language subtitles with a medium effect size. Secondly, the results show that English speakers reading first-language subtitles have significantly higher cognitive load than Chinese speakers reading second-language subtitles with a medium effect size as well. Both of these results could be explained by a lack of habitual behaviour in subtitle reading for English speakers and an unbalanced sample size. Furthermore, comprehension results show that Chinese speakers reading first-language subtitles scored significantly higher than reading second-language subtitles in the laboratory environment.

Chapter 8 Conclusion

8.1 Introduction

This research aims to investigate the impact of the presence or absence of subtitles and the role of subtitle language on audiovisual text processing, cognitive load, and performance in a university context. Based on a quantitative and qualitative analysis of cognitive load, performance, and eye movement data in reading subtitles, the results of the study suggest that the presence of first-language subtitles assists and benefits cognitive processing and thus improves performance. The results indicate that Chinese speakers are more effective in processing first-language subtitles. The findings evidence that reading patterns change with longer mean fixation duration in reading second-language subtitles in a multimedia learning environment. This chapter consists of four sections, section 8.2 presents the benefits of first-language subtitles in educational context, section 8.3 details the limitations of the study and recommendations for future research, and section 8.4 discusses the contribution and possible implications of the results.

8.2 The benefits of L1 subtitles in educational context

The results of this study evidence that reading first-language subtitles is beneficial to second-language learners in a university education context. As online learning is becoming more popular, the presence of subtitles, particularly first-language subtitles, assists second-language learners in comprehending educational content more effectively in a multimedia learning environment that leads to improved performance. Previous studies have proven that the presence of subtitles helped maximising learning (Vanderplank, 1988) and complementing the process of ambiguous or novel information (Bird & William, 2002; Danan, 2004) in the context of second-language learners reading second-language subtitles for the purpose of language acquisition. The current study adds further insight and complement existing research on the effectiveness of subtitles and subtitle language that second-language learners reading first-language subtitles improve performance in an academic context. Instructors and educators can ultilise the impact of first-language subtitles evidenced in this study, allowing teaching materials to be even more accessible that could benefit non-native university students in tertiary education environment. The results provide a valuable implication on education pedagogy in terms of assisting second-language learners in achieving their highest potential academically without being disadvantaged by possible language barriers.

8.3 Limitations and recommendations

8.3.1 Experimental design

The current study investigates the benefits of subtitles in students studying through second-language, and only Chinese speakers are focused on in the study given the posed research questions. Replication of the study involving English speakers or foreign language speakers would provide further information on the role of subtitle language. Furthermore, there is a potential limitation in measuring eye movement patterns between static and dynamic reading, which could be addressed in future studies for further investigation.

8.3.2 Low tracking ratio

Although a high rate of data attrition is common in eye tracking studies, the loss of data in this study is quite high. More accurate eye trackers like the high-speed Eyelink trackers used with a chinrest may result in improved tracking ratio.

8.3.3 Variables in real-life setting

There are many uncontrollable variables in collecting data in real-life situations. As it has happened in this study, these variables range from the number of participants attending on a particular experiment day, to technical issues involving the device used in the experiment, to gaining approval to implement the exact study design in order to answer the research questions.

All these drawbacks or challenges would, in one way or another, have an impact on the power of the study. The answers to minimising the impact of these variables may not be straightforward, but by carefully choosing the sample population in future research, such as recruiting from a course unit that is more prepared and flexible in collaborating research design involving collecting data in a classroom setting, the researcher may have more control in implementing the intended study procedure without compromising the original experimental design.

8.3.4 Styles and topics of the video material

The style of the video material used in this study was paper-hand drawing style. Other formats such as talking head, classroom lecturer and PowerPoint slides presentation could potentially have a different effect on subtitle processing. Future studies could investigate the impact of video styles on subtitle reading in terms of visual complexity and student engagement. Other topic areas could be used in future research to explore the possible impact of topic area in subtitle processing and learning motivation.

8.3.5 Comprehension measurement

The comprehension scores from the classroom setting lacked reliability due to the absence of a tailored comprehension task in the study procedure. The current comprehension scores from the end of term examination involved examining materials that were not covered in the five videos used in the study, therefore the accuracy and validity of what are being tested is questionable. This could be controlled to a much higher degree in the second experiment and future studies should ensure that the comprehension measure can be validated to provide a valid score for accurate analysis which could lead to a more complete interpretation of the research questions. Again, this issue is the result of the study design being compromised. If the researcher has more control in implementing the study procedure, results would be more valid for analysis. Furthermore, another limitation of the study relating to comprehension is that the comprehension data from the laboratory setting mainly tested short-term memory and future research should also look at the long-term effect of first- and second-language subtitles.

8.3.6 Cognitive Load measurement

The accuracy of measuring cognitive load using self-report measures remains subjective. Cognitive load items perceived by Chinese native speakers may be confusing for them to understand. Chinese participants in Experiment 2 constantly asked for clarification on the meaning of the CL items. It seems to suggest that the way each CL item is presented linguistically could be ambiguous, easily confused between rating the difficulty of the content versus the clarity of the instruction.

8.3.7 Sample size

The current study has a sample size of 103 and 70 in Experiment 1 and Experiment 2 respectively, and a sample size of between 12 and 28 in each of the test conditions for both experiments. This sample size may be small compared to empirical reading studies which have relatively larger samples (e.g., Kliegl et al., 2006; Rayner et al., 2011). However, most eye tracking studies in subtitling have relatively small sample sizes, with the number of total sample size usually being under 100 and the sample size for each test condition roughly range between 6 – 30 (e.g., d'Ydewalle & De Bruycker, 2007; Hefer, 2013a, 2013b; Kruger, 2013; Kruger & Steyn, 2013; Perego et al., 2010; Perego et al, 2018; Ross & Knowler, 2013; Szarkowska & Geber-Moron, 2018). In this context, the sample size of the current study provides sufficient power. The difference in these sample sizes depends on the nature of the research and experimental design. Static reading studies tend to have larger samples as the tasks and materials used in these experiments are very simple, such as reading a series of words, sentences or paragraphs, while subtitling studies have more complex, dynamic stimuli and usually involve more than one language, therefore samples are relatively smaller in size.

Regarding other characteristics of sample populations, studies show that age (d'Ydewalle & De Bruycker, 2007; d'Ydewalle et al., 1987) and gender (d'Ydewalle & van Rensbergen, 1989) is not a statistical factor in this type of general language task. Furthermore, the cohort in this study is a representative sample of university students (refer to section 3.3.1 and 5.3).

8.4 Contribution and implications

The current study contributes to the knowledge of a wider research community on the impact of subtitles and subtitle language in first- and second-language learners in a university education context. The results of the current study evidence that first-language subtitles would benefit learners who learn through a second-language medium. The presence of first-language subtitles in academic audiovisual material would potentially assist learners in processing the learning content effectively and thus improve performance. In this study, for Chinese students studying through English as a learning medium, adding first-language subtitles significantly increased their performance. In other words, providing first-language subtitles to Chinese students could have a significant benefit to their academic success and this could also be the case with other language groups. However, adding translated subtitles remains a costly exercise although technologies such as speech recognition software combined with machine translation holds promise. It would be ideal if first-language subtitles can be widely accessible to all students studying through a second-language, and from the results of this study it could be an investment universities across the globe could well consider, particularly in bespoke online courses.

Replication of the current study with improved study design and in different languages and contexts is essential to confirm if results can be replicated. The results of the study confirm the presence of a language barrier that could disadvantage second-language learners in university education even if they have reached the minimum language standard for studying in a second-language country; and it is encouraging that the presence of first-language subtitles could benefit second-language learners in an academic context. Future research would also benefit from carefully controlling the participants' language proficiency and their reading habits in L2. This would allow further investigation to verify if the advantage for reading L1 subtitles decreases for those participants with advanced or near-native listening and reading skills in L2.

Furthermore, the results show that there is a significant difference in eye movement patterns between reading first- and second-language text dynamically while there is no significant difference in eye movement patterns between reading first- and second-language text statically. These results have important implications for pedagogy in supporting academic success for second-language learners.

References

Akaike, H. (1974). A new look at the statisticalmodel identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225-231.

Antonenko, P. D., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior, 26*(2), 140-150. doi:https://doi.org/10.1016/j.chb.2009.10.014

Ayres, P. (2006a). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*(3), 287-298. doi:10.1002/acp.1245

Ayres, P. (2006b). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and instruction, 16*(5), 389-400. doi:https://doi.org/10.1016/j.learninstruc.2006.09.001

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

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), 136-140. https://doi-org.simsrad.net.ocs.mq.edu.au/10.1016/j.cub.2009.12.014

Baddeley, A. (2012). Working memory: Theories, models, and controversiers. *Annual Review of Psychology, 63*(1), 1-29.

Baltova, I. (1999). Multisensory Language Teaching in a Multidimensional Curriculum: The Use of Authentic Bimodal. *Canadian Modern Language Review, 56*(1), 31. doi:https://doi.org/10.3138/cmlr.56.1.31

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

BeGaze™ [Computer software]. (2016). SensoMotoric Instruments GmbH. Teltow, Germany.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(04), 509-533. doi:doi:10.1017/S0142716402004022

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(02), 399-418. doi:doi:10.1017/S0142716412000434

Bochkarev, V. V., Shevlyakova, A. V., & Solovyev, V. D. (2015). The average word length dynamics as an indicator of cultural changes in society. *Social Evolution & History, 14*(2), 153-175.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal, 9*(2), 378-400.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*, 261-304. doi:10.1177/0049124104268644

Chan, W. S., Kruger, J-L., & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 18*, 237-272.

Chang, C., & Chen, Y. (2002). The western computer and the Chinese charater: Recent debates on recent Chinese writing reform. *Intercultural Communication Studies XI, 3*, 67-84.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. USA: Lawrence Erlbaum Associates.

Crawley, M. J. (2007). *The R Book*. England: John Wiley & Sons.

Danan, M. (1992). Reversed Subtitling and Dual Coding Theory: New Directions for Foreign Language Instruction. *Language Learning, 42*(4), 497-527.

Danan, M. (2004). Captioning and Subtitling: Undervalued Language Learning Strategies. *Meta: Journal des traducteurs, 49*(1), 67-77. doi:10.7202/009021ar

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*(2), 105-134. doi:10.1007/s11251-009-9110-0

Debue, N., & van de Leemput, C. (2014). What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology, 5*(1099). doi:10.3389/fpsyg.2014.01099

Dehaene, S. (2009). *Reading in the brain: the science and evolution of a human invention*. New York: Viking.

Diao, Y., Chandler, P., & Sweller, J. (2007). The Effect of Written Text on Comprehension of Spoken English as a Foreign Language. *The American Journal of Psychology, 120*(2), 237-261.

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and instruction, 17*(1), 78-88. doi:http://dx.doi.org/10.1016/j.learninstruc.2006.11.007

Diaz-Cintas, J., & Remael, A. (2007). *Audiovisual Translation: Subtitling.* Kinderhook, NY: St. Jerome Pub.

Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In T. Dwyer, Perkins, C., Redmond, S., & Sita, J. (Ed.), *Seeing into Screens: Eye Tracking and the Moving Image* (pp. 46-64). London: Bloomsbury.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye Movements of Children and Adults While Reading Television Subtitles. *European Psychologist, 12*(3), 196-205.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching Subtitled Television: Automatic Reading Behavior. *Communication Research, 18*(5), 650-666. doi:10.1177/009365091018005005

d’Ydewalle, G., & van Rensbergen, J. (1989). Developmental Studies of Text-Picture Interactions in the Perception of Animated Cartoons with Text. In H. Mandl and J. R. Levin (Eds.), *Knowledge Acquisition from Text and Pictures*, (pp. 233–248). Amsterdam: Elsevier Science. doi: 10.1016/S0166-4115(08)62157-3

d'Ydewalle, G., van Rensbergen, J., & Pollet, J., (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J. L. O’Regan and A. Levy-Schoen (Eds.), *Eye movements: From Physiology to Cognition* (pp. 313-321). North Holland: Elsevier Science Publishers.

Experiment Center™ [Computer software]. (2010). SensoMotoric Instruments GmbH. Teltow, Germany.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861-874.

Feng, G., Muller, K., Shu, H., & Zhang, H. (2009). Orthography and the development of reading processes: An eye-movement study of Chinese and English. Child Development, 80(3), 720-735.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32* (3), 221-233.

Flesch, R. (1979). *How to write plain English*. New York: Harper & Row, Publishers.

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression (3rd ed). Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frasca, R. R. (2007). *Microeconomics.* Retrieved from https://academic.udayton.edu/pmic/Instructors%20manual_copy(1).htm

Garza, T. J. (1991). Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning. *Foreign Language Annals, 24*(3), 239-258. doi:10.1111/j.1944-9720.1991.tb00469.x

Geary, D. C. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1-99). Greenwich, CT: Information Age Publishing.

Geary, D. C. (2008). An Evolutionarily Informed Education Science. *Educational Psychologist, 43*(4), 179-195. doi:10.1080/00461520802392133

generated and corrected subtitles on cognitive load and learning in a first- and second-

Gernsbacher, M. A. (2015). Video Captions Benefit Everyone. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 195-202. doi:10.1177/2372732215602130

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and instruction, 15*(4), 313-331. doi:https://doi.org/10.1016/j.learninstruc.2005.07.001

Greco, G. M. (2018). The nature of accessibility studies. *Journal of Audiovisual Translation, 1*(1), 205-232.

Greco, G. M. (2019). Accessibility studies: Abuses, misuses and the method of poietic design. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11786*, 15-27. https://doi-org.simsrad.net.ocs.mq.edu.au/10.1007/978-3-030-30033-3_2

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36,* 85-93.

Guo, P. J., Kim, J., & Rubin, R. (2014). *How video production affects student engagement: an empirical study of MOOC videos.* Paper presented at the Proceedings of the first ACM conference on Learning@ scale conference.

Hartig, F. (2019). DHARMa: *Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.2.5. https://CRAN.R-project.org/package=DHARMa

Hayati, A., & Mohmedi, F. (2009). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181-192. doi:10.1111/j.1467-8535.2009.01004.x

Hefer, E. (2013a). Reading first and second language subtitles: Sesotho viewers reading in Sesotho and English. *Southern African Linguistics and Applied Language Studies, 31*(3), 359-373. doi:10.2989/16073614.2013.837610

Hefer, E. (2013b). Reading second language subtitles: A case study of Afrikaans viewers reading in Afrikaans and English. *Perspectives, 21*(1), 22-41. doi:10.1080/0907676X.2012.722652

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. v. d. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*(3), 786-797. doi:http://dx.doi.org/10.1016/j.chb.2007.02.009

Ivarsson, J., & Carroll, M. (1998). *Subtitling*. Simrishamn, Sweden: TransEdit HB.

iViewETG [Computer software]. (2016). SensoMotoric Instruments GmbH. Teltow, Germany.

iViewX™ [Computer software]. (2016). SensoMotoric Instruments GmbH. Teltow, Germany.

Jakobson, R. (1959). On linguistic aspects of translation. In A. Fang & R. A. Brower (Eds.), *On translation* (Vol. 23, pp. 232-239). Boston: Harvard University Press.

Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American Annals of the Deaf, 145*(3), 275-285.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441-480. doi:http://dx.doi.org/10.1016/0010-0285(76)90015-3

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19,* 402-410.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38* (1), 23-31. DOI: 10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied cognitive psychology, 13*(4), 351-371.

Kalyuga, S., & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 247-262). Cambridge: Cambridge University Press.

Karunaratne, P. (2012). *Demand & supply and government actions in markets* [Video]. Retrieved from https://www.youtube.com/watch?v=Z3Qb94pubAQ

Karunaratne, P. (2015a). *Market structure and efficiency* [Video]. Retrieved from https://www.youtube.com/watch?v=5pxkCxMYUPU

Karunaratne, P. (2015b). *Perfect competition and monopolistic competition* [Video]. Retrieved from https://www.youtube.com/watch?v=3T9GSJeomKE

Karunaratne, P. (2015c). *Producer theory and shifting cost curves* [Video]. Retrieved from https://www.youtube.com/watch?v=C3PDVmPijIY

Karunaratne, P. (2015d). *Externalities* [Video]. Retrieved from https://www.youtube.com/watch?v=GE8XeeaVkJ8

Kassambara, A. (2019). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.2.4. https://CRAN.R-project.org/package=ggpubr

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report, 8-75. *Institute for Simulation and Training. 56.*

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimantal Psychology: General, 135*(1), 12-35. doi: 10.1037/0096-3445.135.1.12

Koda, K. (1990). The Use of L1 Reading Strategies in L2 Reading: Effects of L1 Orthographic Structures on L2 Phonological Recoding Strategies. *Studies in Second Language Acquisition, 12*(4), 393-410. doi:10.1017/S0272263100009499

Koda, K. (1994). Second language reading research: Problems and possibilities. *Applied Psycholinguistics, 15*(1), 1-28. doi:10.1017/S0142716400006950

Kruger, J.-L. (2016). Psycholinguistics and audiovisual translation. *Target, 28*(2), 276-287. doi:10.1075/target.28.2.08kru

Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32*(6), 19-31. doi:doi.org/10.14742/ajet.3084

Kruger, J.-L., Hefer, E., & Matthew, G. (2013). Measuring the Impact of Subtitles on Cognitive Load: Eye Tracking and Dynamic Audiovisual Texts. *Proceeding of Eye Tracking South Africa, Cape Town*, 62-66.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs L2. *Journal of Eye Movement Research, 7*(5), 1-15.

Kruger, J.-L., & Steyn, F. (2013). Subtitles and eye tracking reading and performance. *Reading Research Quarterly, 49*(1), 105-120. doi:10.1002/rrq.59

Kuo, S. (2014). *Quality in subtitling: theory and professional reality.* (Doctoral dissertation), Imperial College London, London, UK. Retrieved from http://simsrad.net.ocs.mq.edu.au/login?url=https://search-proquest-com.simsrad.net.ocs.mq.edu.au/docview/1785480858?accountid=12219

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26. doi:10.18637/jss.v082.i13

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? *International Journal of Psychology, 46*(6), 455-462. doi:10.1080/00207594.2011.565343

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied cognitive psychology, 25*(6), 943-951. doi:10.1002/acp.1787

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44*(1), 107-123. doi:10.1007/s11251-015-9362-9

Leppink, J., Paas, F., van der Vleuten, C. P., van Gog, T., & van Merrienboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavioural Research Methods, 45*(4), 1058-1072. doi:10.3758/s13428-013-0334-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and instruction, 30*, 32-42. doi:10.1016/j.learninstruc.2013.12.001

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119-127. doi:10.1007/s40037-015-0192-x

Liao, S., Kruger, J-L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation, 33*, 70-98.

Lincoln, F., & Rademacher, B. (2006). Learning Styles of ESL Students in Community Colleges. *Community College Journal of Research and Practice, 30*(5-6), 485-500. doi:10.1080/10668920500207965

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55-75). UK: Elsevier.

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 227-246). New York: Cambridge University Press.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940-967. doi:doi:10.1044/1092-4388(2007/067)

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321-328. doi:10.1111/j.1944-9720.1999.tb01344.x

Markham, P., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439-445. doi:doi:10.1111/j.1944-9720.2001.tb02083.x

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. *New Directions for Teaching & Learning, 2002*(89), 55-71.

Mayer, R. E., & Fiorella, L. (2014). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 279-315). New York: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: when presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia Learning in a Second Language: A Cognitive Load Perspective. *Applied Cognitive Psychology, 28*(5), 653-660. doi:10.1002/acp.3050

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638-643.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words?: extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist, 50*(1), 31-42. doi:10.1080/00461520.2015.1004068

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLOS ONE, 4*(11), e7785. doi:10.1371/journal.pone.0007785

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720-739. doi:https://doi.org/10.1016/j.system.2013.07.013

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163.

Orero, P., Doherty, S., Kruger, J-L., Matamala, A., Pedersen, J., Perego, E., Romero-Fesco, P., Rovira-Esteva, S., Soler-Vilageliu, O., & Szarkowska, A. (2018). Conducting experimental research in audiovisual translation (AVT): A position paper. *The Journal of Specialised Translation, 30*, 105-126.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27-45. doi:10.1007/s10648-011-9179-2

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 27-42). New York: Cambridge University Press.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71. doi:10.1207/s15326985ep3801_8

Paas, F., & van Merrienboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351.

Paivio, A. (1990). *Mental Representation: A Dual Coding Approach.* New York: Oxford University Press.

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255-287.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The Cognitive Effectiveness of Subtitle Processing. *Media Psychology, 13*(3), 243-272. doi:10.1080/15213269.2010.502873

Perego, E., Del Missier, F., & Straga, M. (2018). Dubbing vs. subtitling complexity matters. *Target-International Journal of Translation Studies, 30*(1), 137-157. doi:10.1075/target.16083.per

R Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R project.org/

Rajendran, D. J., Duchowski, A. T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives, 21*(1), 5-21. doi:10.1080/0907676X.2012.722651

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K. (2004). Eye movements, cognitive processes, and reading. Studies of Psychology and Behavior, 2, 482–488.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506. doi:10.1080/17470210902816461

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(3), 241-255.

Rayner, K., Juhasz, B., & Pollatsek, A. (2005). Eye movements during reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 79-97). MA USA: Blackwell Publishing.

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability of the eye movements of Chinese readers. Psychonomic Bulletin & Review, 12(6), 1089-1093.

Rayner, K., & Reingold, E. M. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences, 1*, 107-112. doi:http://dx.doi.org/10.1016/j.cobeha.2014.10.008

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77. DOI: 10.1186/1471-2105-12-77

Romero-Fresco, P. (2015). Final thoughts: viewing speed in subtitling. In P. Romero-Fresco, (Ed.), *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe* (pp. 335–342). Berlin: Peter Lang.

Romero-Fresco, P. (2018). Eye tracking, subtitling and accessible filmmaking. In T. Dwyer, Perkins, C., Redmond, S., & Sita, J. (Ed.), *Seeing into Screens: Eye Tracking and the Moving Image* (pp. 253-258). London: Bloomsbury.

Ross, N. M., & Kowler, E. (2013). Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision, 13*(4), 1-1. doi:10.1167/13.4.1

Schotter, E., & Rayner, K. (2012). Eye movements in reading: Implication for reading subtitles. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 83-104). Rome: Aracne Editrice.

Shaw, R. G., & Michell-Olds, T. (1993). ANOVA for unbalanced data: an overview. *Ecology, 74*(6), 1638-1645.

SMI (SensoMotoric Instruments). (2019). BeGaze™ manual: version 3.7. Teltow: SMI.
SMI ETG [Apparatus]. (2016). SensoMotoric Instruments GmbH. Teltow, Germany.
SMI RED250™ [Apparatus]. (2011). SensoMotoric Instruments GmbH. Teltow, Germany.
SMI Semantic Gaze Mapping [Computer software]. (2016). SensoMotoric Instruments GmbH. Teltow, Germany.
Specker, E. (2015). *L1/L2 Eye movement reading of closed captioning: A multimodal analysis of multimodal use.* (PhD), The University of Arizona, Retrieved from http://hdl.handle.net/10150/194820
Sung, Y-T., Chang, T-H., Lin, W-C., Hsieh, K-S. (2016). CRIE: An automated analyzer for Chinese texts. *Behavior Research Methods, 48*(4), 1238-1251. DOI:10.3758/s13428-015-0649-1
Sun, F., & Feng, D. (1999). Eye movements in reading Chinese and English text. In J. Wang, H.-C. Chen, R. Radach, & A. Inhoff (Eds.), *Reading Chinese Script: A Cognitive Analysis* (pp. 189-206). New York: Taylor and Francis.
Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics, 37*, 502-506. doi:10.3758/bf03204913
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. doi:http://dx.doi.org/10.1016/0364-0213(88)90023-7
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.
Sweller, J. (2003). Evolution of human cognitive architecture. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 43, pp. 215-266). USA: Elsevier Science.
Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*, 9-31.
Sweller, J. (2006). How the human cognitive system deal with complexity. In J. Elen & R. E. Clark (Eds.), *Handling Complexity in Learning Environments: Theory and Research* (pp. 13-26). The Netherlands: Elsevier.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138. doi:10.1007/s10648-010-9128-5
Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37-76). USA: Elsevier Science.
Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.
Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*(1), 434-458. doi:10.1177/147470490600400135
Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.
Sydorenko, T. (2010). Modality of input and vocabulart acquisition. Language Learning & Technology, 14 (2), 50-73.
Szarkowska, A., & Geber-Moron, O. (2018). Viwers can keep up with fast subtitles: Evidence form eye movements. *PLOS ONE, 13*(6), e0199331. doi:https://doi.org/10.1371/journal.pone.0199331
Szarkowska, A., Krejtz, I., Klyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf, 156*(4), 363-378.
Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, Ł., & Kruger, J.-L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers.

Across Languages and Cultures, 17(2), 183-204. doi:10.1556/084.2016.17.2.3

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71-81.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422-427. doi:10.1111/j.1944-9720.2005.tb02228.x

Tsai, C.-J., & Huang, S. C. (2009). *Target language subtitles for comprehensible file language input.* doi:http://120.107.180.177/1832/9901/099-1-11p.pdf

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from online education videos. *Journal of Media Psychology Theories Methods and Applications, 29*, 18-30. doi:10.1027/1864-1105/a000208

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal, 42*(4), 272-281. doi:10.1093/elt/42.4.272

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System, 18*(2), 221-234. doi:http://dx.doi.org/10.1016/0346-251X(90)90056-B

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching:subtitles forthe deaf and hard-of-hearing as tools for language learning*. London: Palgrave Macmillian.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.

Wickham, H., & Bryan, J. (2019). *readxl: Read Excel Files.* R package version 1.3.1. https://CRAN.R-project.org/packagereadxl

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A Grammar of Data Manipulation.* R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign langauge listening activities. *Language Learning & Technology, 14*(1), 65-86. doi:http://dx.doi.org/10125/44203

Winke, P., Sydorenko, T., & Gass, S. (2013). Factors influencing the use of captions by foreign language learners: an eye-tracking study. *The Modern Language Journal, 97*(1), 254-275

Winters, R. (2017). *Practical predictive analytics: Back to the future with R, Spark, and more!* Birmingham, England: Packt Publishing.

Appendix A

Cognitive load test on extraneous load adapted from Leppink et al. (2014)

Student ID: ________________________

All of the following questions refer to the video that just finished. Please take your time to read each of the questions carefully and respond (by placing a tick ✓) to each of the questions on the presented scale from 0 to 10, in which ‘0’ indicates *not at all the case* and ‘10’ indicates *completely the case*:

1. The explanations and instructions in this video were very unclear.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										**Completely the case**

2. The explanations and instructions in this video were full of unclear language.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										**Completely the case**

3. The explanations and instructions in this video were, in terms of learning, very ineffective.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										**Completely the case**

4. I invested a very high mental effort in unclear and ineffective explanations and instructions in this video.

0	1	2	3	4	5	6	7	8	9	10
Not at all the case										**Completely the case**

Appendix B

Multiple choice questions and answers for the five videos (Frasca, 2007)

Video 1. Demand & Supply and Government action in market

1[16]. What is the best output for society to produce as a whole?

A) when marginal benefit is larger than marginal cost.

B) when marginal benefit is smaller than marginal cost

C) when marginal cost is larger than marginal benefit

D) when marginal benefit equals marginal cost.

2[17]. Which of the following is a typical effect of a price ceiling set below the equilibrium price?

A) People can buy more than they can at the equilibrium price because the ceiling price is lower.

B) The price ceiling has no effect on the market equilibrium.

C) Less of the good is produced with the ceiling than would be produced without the ceiling.

D) None of the above answers are correct.

3[18]. Which of the following is the best way to describe equilibrium in a market? At equilibrium,

A) the price charged is usually affordable to most people.

B) the supply and demand curves can never shift again.

C) the quantity supplied equals the quantity demanded.

D) the price charged is the lowest possible.

---

[16] Question created by researcher.
[17] Question chose from Q33 http://academic.udayton.edu/pmic/Answers/micro%20final%20%202006.pdf
[18] Question chose from Q18 http://academic.udayton.edu/pmic/Quizzes/micro%20quiz%202.pdf

4[19]. If there are no external costs no external benefits, with no government intervention,

A) Consumer surplus is larger than producer surplus.

B) Consumer surplus is smaller than producer surplus.

C) The society is not allocatively efficient.

D) The society has the largest total surplus combination of consumer and producer.

5[20]. Deadweight loss is the decrease in ________ from producing an inefficient amount of a product.

A) profit

B) consumer surplus

C) producer surplus

D) consumer surplus and producer surplus

Video 2. Externalities

1[21]. An externality is a cost or a benefit from an economic transaction that falls on

A) people who did not participate in the transaction.

B) consumers of the good but not producers.

C) producers of the good but not consumers.

D) both consumers and producers of the good.

2[22]. An externality can be a

A) marginal cost but not a total cost.

B) benefit but not a cost.

C) cost or a benefit.

---

[19] Question chose from Q48 http://academic.udayton.edu/pmic/MC%20Questions/Chap%205%20mc.pdf
[20] Question chose from Q96 http://academic.udayton.edu/pmic/MC%20Questions/Chap%205%20mc.pdf
[21] Question chose from Q3 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2014%20testbank.pdf
[22] Question chose from Q2 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2014%20testbank.pdf

D) cost but not a benefit

3[23]. Market failures can result from

A) external benefits and external costs.

B) neither external benefits nor external costs.

C) external costs but not external benefits.

D) external benefits but not external costs.

4[24]. Which of the following IS a negative externality created by consumers?

A) Green technology

B) Eating an apple

C) Smoking

D) All of the above

5[25]. Which of the following is an example of positive externality?

A) Smoking.

B) Vaccination.

C) Pollution.

D) All of the above are correct.

Video 3. Producer Theory and shifting cost curves

1[26]. In the short run,

A) there are no variable costs.

[23] Question chose from http://academic.udayton.edu/pmic/MC%20Questions/Chap%2014%20testbank.pdf
[24] Question created by researcher.
[25] Question created by researcher.
[26] Question chose from Q1 http://academic.udayton.edu/pmic/Answers/micro%20quiz%204.pdf

B) at least one resource is fixed.

C) all resources are variable.

D) all resources are fixed.

2[27]. An example of a variable resource in the short run is

A) an employee.

B) capital equipment.

C) land.

D) a building.

3[28]. As output increases, average fixed cost

A) decreases continuously.

B) decreases, then increases.

C) remains constant.

D) increases, then decreases.

4[29]. Total cost is

A) the difference between the average variable cost and the average fixed cost.

B) the sum of the total fixed cost and the total variable cost.

C) the product of the marginal cost times the average total cost.

D) None of the above.

5[30]. The law of diminishing returns makes it clear that as more a variable input is employed,

[27] Question chose from Q4 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2010%20MC.pdf
[28] Question chose from Q13 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2010%20MC.pdf
[29] Question chose from Q43 http://academic.udayton.edu/pmic/Answers/micro%20final%20%202006.pdf
[30] Question chose from Q10 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2010%20MC.pdf

A) in the long-run the marginal product of the variable input will eventually fall.

B) in the short-run the marginal product of the variable input will eventually fall.

C) in the short-run the marginal product of the variable input will eventually rise.

D) in the long-run the marginal product of the variable input will eventually rise.

Video 4. Market Structures and Efficiency

1[31]. In perfect competition, restrictions on entry into an industry

A) do not exist.

B) apply to labor but not to capital.

C) apply to both capital and labor.

D) apply to capital but not to labor.

2[32]. Monopolistic competition is a market structure in which

A) there are barriers to entry.

B) a small number of firms compete.

C) firms only compete on product price.

D) each firm produces a differentiated product.

3[33]. The concept of elasticity of supply measures the responsiveness of the

A) quantity supplied to a change in price.

B) quantity demanded to a change in quantity supplied.

C) price to a change in quantity supplied.

D) quantity supplied to a change in quantity demanded.

---

[31] Question chose from Q4 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2011%20MC.pdf
[32] Question chose from Q77 http://academic.udayton.edu/pmic/Answers/micro%20final%20%202006.pdf
[33] Question chose from Q21 http://academic.udayton.edu/pmic/Answers/Mba%20640%202.pdf

4[34]. An oligopoly is a market structure in which there are

A) many sellers selling a differentiated product.

B) only a few buyers but many sellers.

C) a few products sold by many sellers.

D) only a few sellers selling either an identical or differentiated product with barriers to entry.

5[35]. Perfect competition is an industry with

A) a few firms producing identical goods.

B) many firms producing goods that differ somewhat.

C) a few firms producing goods that differ somewhat in quality.

D) many firms producing identical goods.

Video 5. Perfect competition and monopolistic competition

1[36]. A characteristic of monopolistic competition is

A) a low ratio of fixed to variable costs.

B) a high capital-output ratio.

C) product differentiation.

D) the absence of advertising.

2[37]. In the long run, in monopolistic competition

A) firms earn an economic profit.

B) firms earn zero economic profit.

C) price equals marginal cost.

[34] Question chose from Q89 http://academic.udayton.edu/pmic/Answers/micro%20final%20%202006.pdf
[35] Question chose from Q1 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2011%20MC.pdf
[36] Question chose from Q5 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2013%20MC.pdf
[37] Question chose from Q87 http://academic.udayton.edu/pmic/Answers/micro%20final%20%202006.pdf

D) Both answers A and C are correct

3[38]. For a firm in monopolistic competition, the marginal cost curve intersects the average total cost curve

A) at no point.

B) at the minimum average total cost.

C) to the left of the minimum average total cost.

D) to the right of the minimum average total cost.

4[39]. Which of the following is NOT considered excess capacity?

A) Spare hotel rooms.

B) Spare captial.

C) Spare tables in a restaurant.

D) All of the above are correct.

5[40]. A monopolistically competitive firm has _______ power to set the price of its product because ________.

A) no; there are no barriers to entry

B) some; there are barriers to entry

C) some; of product differentiation

D) no; of product differentiation

---

[38] Question chose from Q17 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2013%20MC.pdf
[39] Question created by researcher.
[40] Question chose from Q1 http://academic.udayton.edu/pmic/MC%20Questions/Chap%2013%20MC.pdf

Answers

	Video 1	**Video 2**	**Video 3**	**Video 4**	**Video 5**
Q1	D	A	B	A	C
Q2	C	C	A	D	B
Q3	C	A	A	A	B
Q4	D	C	B	D	B
Q5	D	B	B	D	C

Appendix C

Student number: ________________________________

Biographical questionnaire

We thank you for participating in this study. Your contribution is valuable for assisting current knowledge of processes related to the impact of subtitles in learning.

We would like you to know that this test only has academic objectives. The information that you provide will be used only within scientific domains.

This empirical experience has three parts. During the first part, we will ask you to give us some information about you and your language background. After this, you will watch a video. Finally, we will present you with another 4-items questionnaire related to the how difficult you feel about the video presented. Now, to begin with, answer these questions.

1. In what language did you complete your secondary schooling?

____

2. What is your gender? ______________________________________

3. What is your date of birth? __________________

3. What is your home country? __________________

4. Indicate how many months / years you have spent in an English-speaking country in the last 10 years. _______________________

Appendix D

Chapter 4

```
>
> CL_class$Group <- factor(CL_class$Group, levels = c("No_sub", "Chinese_sub", "English_sub"))
> CL_anova <- aov(CL ~ Language * Group, data = CL_class)
> Anova(CL_anova, type = "III")
Anova Table (Type III tests)

Response: CL
                Sum Sq  Df  F value    Pr(>F)
(Intercept)     8112.3   1 133.2695 < 2.2e-16 ***
Language          74.9   1   1.2301  0.268057
Group            135.1   2   1.1094  0.330760
Language:Group   797.0   2   6.5465  0.001595 **
Residuals      24165.9 397
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Figure 4-1. ANOVA output for CL classroom study.

```
>
> TukeyHSD(CL_anova, which = "Language:Group")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = CL ~ Language * Group, data = CL_class)

$`Language:Group`
                                              diff        lwr       upr     p adj
English:No_sub-Chinese:No_sub            1.6951754 -2.6817535 6.0721044 0.8775159
Chinese:Chinese_sub-Chinese:No_sub      -0.3600571 -4.1761089 3.4559947 0.9998046
English:Chinese_sub-Chinese:No_sub      -2.4631579 -6.5956570 1.6693413 0.5280030
Chinese:English_sub-Chinese:No_sub      -1.6894399 -5.3715074 1.9926275 0.7772606
English:English_sub-Chinese:No_sub       3.0285088 -1.3484202 7.4054377 0.3548121
Chinese:Chinese_sub-English:No_sub      -2.0552326 -6.0806919 1.9702268 0.6886557
English:Chinese_sub-English:No_sub      -4.1583333 -8.4849511 0.1682844 0.0675879
Chinese:English_sub-English:No_sub      -3.3846154 -7.2832937 0.5140629 0.1306722
English:English_sub-English:No_sub       1.3333333 -3.2273222 5.8939889 0.9604067
English:Chinese_sub-Chinese:Chinese_sub -2.1031008 -5.8613404 1.6551389 0.5973717
Chinese:English_sub-Chinese:Chinese_sub -1.3293828 -4.5858282 1.9270626 0.8513525
English:English_sub-Chinese:Chinese_sub  3.3885659 -0.6368935 7.4140253 0.1550055
Chinese:English_sub-English:Chinese_sub  0.7737179 -2.8483995 4.3958354 0.9901631
English:English_sub-English:Chinese_sub  5.4916667  1.1650489 9.8182844 0.0042322
English:English_sub-Chinese:English_sub  4.7179487  0.8192704 8.6166270 0.0076878
```

Figure 4-2. TukeyHSD output between the means of groups for CL classroom study.

```
> CL_class$Lang3 <- factor(CL_class$Lang3, levels = c("English", "Others", "Chinese"))
> CL_anova <- aov(CL ~ Lang3 * Group, data = CL_class)
> Anova(CL_anova, type = "III")
Anova Table (Type III tests)

Response: CL
             Sum Sq  Df F value    Pr(>F)    
(Intercept)    4302   1 59.9319 3.765e-14 ***
Lang3           132   2  0.9174   0.40008    
Group           567   2  3.9468   0.01978 *  
Lang3:Group     792   4  2.7597   0.02698 *  
Residuals     46726 651                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Figure 4-4. ANOVA output for CL between Chinese, English and foreign speakers.

```
> kruskal.test(Comp ~ Group, data = Comp_class)

        Kruskal-Wallis rank sum test

data:  Comp by Group
Kruskal-Wallis chi-squared = 1.691, df = 2, p-value = 0.4294

>
>
>
>
>
>
>
> kruskal.test(Comp ~ Language, data = Comp_class)

        Kruskal-Wallis rank sum test

data:  Comp by Language
Kruskal-Wallis chi-squared = 7.4843, df = 1, p-value = 0.006224
```

Figure 4-6. Kriskal-Wallis rank sum test output summary for comprehension.

Chapter 5

```
quantile(Fix_MW$Fix_MW, 0.95, na.rm=TRUE)
quantile(Fix_MW$Fix_MW, 0.99, na.rm=TRUE)

# since the 99% quantile is 1.76, by creating a subset smaller than the 99% quantile
# can exclude the outliner

Fix_MW_clean <- subset(Fix_MW, Fix_MW < quantile(Fix_MW, 0.99, na.rm=TRUE))
hist(Fix_MW_clean$Fix_MW)
qqnorm(Fix_MW_clean$Fix_MW)


#Now we fit a model that distinguishes between zero and non0zero using a logistical regression to look at the effect of the different variables

#Because of all the zero values we need to explore a binomial model that looks at zero and non-zero data separately.

# need another variable in dataset to identify zeros and non-zeros that are unprocoessed or processed.

Fix_MW_clean$processed <- Fix_MW_clean$FC > 0

#centered CPS as average speed for each subtitle langauge so the intecept does not base on zero CPS
Fix_MW_clean$CPS_centred <- NA
Fix_MW_clean$CPS_centred[Fix_MW_clean$Group == "Chinese_sub"] <- scale(Fix_MW_clean$CPS[Fix_MW_clean$Group == "Chinese_sub"])
Fix_MW_clean$CPS_centred[Fix_MW_clean$Group == "English_sub"] <- scale(Fix_MW_clean$CPS[Fix_MW_clean$Group == "English_sub"])

tapply(Fix_MW_clean$CPS_centred, Fix_MW_clean$Group, mean)
lang_meanCPS <- tapply(Fix_MW_clean$CPS_centred, Fix_MW_clean$Group, mean)

# distribution family change to poisson as Fix_MW is counts not continuous
# Mean word has been centered as the length of subtitle could affect the number of fixation count.

Fix_MW_clean$MW_centered <- scale(Fix_MW_clean$MW)

View(Fix_MW_clean)
```

Figure 5-3. Data preparation for computation by scaling and cleaning in R.

```
>
> CL_aov <- aov(CL ~ Group, data = CL)
> # Summary of the analysis
> summary(CL_aov)
             Df Sum Sq Mean Sq F value Pr(>F)
Group         2     42   20.91    0.28  0.756
Residuals   347  25866   74.54
>
>
>
> CL_aov1 <- aov(CL ~ Group*SL, data = CL)
> summary(CL_aov1)
             Df Sum Sq Mean Sq F value Pr(>F)
Group         2     42   20.91   0.280  0.756
SL            1    110  109.71   1.467  0.227
Group:SL      2     33   16.43   0.220  0.803
Residuals   344  25724   74.78
>
>
>
> CL_aov2 <- aov(CL ~ Group + SL, data = CL)
> summary(CL_aov2)
             Df Sum Sq Mean Sq F value Pr(>F)
Group         2     42   20.91   0.281  0.755
SL            1    110  109.71   1.474  0.226
Residuals   346  25757   74.44
>
```

Figure 6-1. ANOVA computation showing SL has no significant impact on CL.

```
>
> CL_aov <- aov(CL ~ Group, data = CL)
> # Summary of the analysis
> summary(CL_aov)
             Df Sum Sq Mean Sq F value Pr(>F)
Group         2     42   20.91    0.28  0.756
Residuals   347  25866   74.54
>
```

Figure 6-2. ANOVA output for CL ratings.

```
> kruskal.test(Comp ~ Group, data = Comp_lab)

        Kruskal-Wallis rank sum test

data:  Comp by Group
Kruskal-Wallis chi-squared = 7.0096, df = 2, p-value = 0.03005

> #From the output of the Kruskal-Wallis test,we know that there is a significant difference between groups
> #but we don't know which pairs of groups are different.
> #wilcox text is used to calculate pairwise comparisons between group levels with corrections for multiple testing.
> pairwise.wilcox.test(Comp_lab$Comp, Comp_lab$Group,
+                      p.adjust.method = "BH")

        Pairwise comparisons using Wilcoxon rank sum test

data:  Comp_lab$Comp and Comp_lab$Group

            Chinese_sub English_sub
English_sub 0.042       -
No_sub      0.042       0.823

P value adjustment method: BH
>
```

Figure 6-4. Kruskal-Wallis rank sum test for comprehension in R.

```
>
> plot(roc(Fix_MW_clean$processed, fitted.values(fit_Fix_MW_binomial)))
Setting levels: control = FALSE, case = TRUE
Setting direction: controls < cases
> roc(Fix_MW_clean$processed, fitted.values(fit_Fix_MW_binomial))
Setting levels: control = FALSE, case = TRUE
Setting direction: controls < cases

Call:
roc.default(response = Fix_MW_clean$processed, predictor = fitted.values(fit_Fix_MW_binomial))

Data: fitted.values(fit_Fix_MW_binomial) in 2134 controls (Fix_MW_clean$processed FALSE) < 10693 cases (Fix_MW_clean$processed TRUE).
Area under the curve: 0.8626
```

Figure 6-7. ROC curve computation of the binomial model in R.

```
>
>
>
> fit_Fix_MW_binomial <- glmer(processed ~ Group +  (1|Participant) + (1|video/AOI), family=binomial(), data=Fix_MW_clean)
> summary(fit_Fix_MW_binomial)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: processed ~ Group + (1 | Participant) + (1 | video/AOI)
   Data: Fix_MW_clean

     AIC      BIC   logLik deviance df.resid
  9028.4   9065.7  -4509.2   9018.4    12822

Scaled residuals:
    Min      1Q  Median      3Q     Max
-9.3668  0.1019  0.2482  0.4097  4.0738

Random effects:
 Groups      Name        Variance Std.Dev.
 AOI:video   (Intercept) 0.721250 0.84926
 Participant (Intercept) 2.628404 1.62124
 video       (Intercept) 0.004983 0.07059
Number of obs: 12827, groups:  AOI:video, 403; Participant, 45; video, 5

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.6692     0.3529   4.730 2.24e-06 ***
GroupEnglish_sub   0.8222     0.4912   1.674   0.0942 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr)
GrpEnglsh_s -0.700
>
> exp(fixef(fit_Fix_MW_binomial)["GroupEnglish_sub"])
GroupEnglish_sub
        2.275532
>
```

Figure 6-9. glmer binomial computation in R.

```
>
>
> # Correlation between skipped English subtitles and NRN
> skipped_Eng = c(12.52, 10.36, 12.52, 17.24, 13.21)
> NRN = c(37.04, 53.61, 45.56, 19.18, 34.43)
> cor.test(skipped_Eng,NRN)

        Pearson's product-moment correlation

data:  skipped_Eng and NRN
t = -5.945, df = 3, p-value = 0.009516
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9974556 -0.5087638
sample estimates:
       cor
-0.9600826

> # Correlation between skipped English subtitles and RN
> skipped_Eng = c(12.52, 10.36, 12.52, 17.24, 13.21)
> RN = c(7.41, 4.12, 12.22, 28.77, 13.11)
> cor.test(skipped_Eng, RN)

        Pearson's product-moment correlation

data:  skipped_Eng and RN
t = 8.1554, df = 3, p-value = 0.003856
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7002055 0.9986212
sample estimates:
      cor
0.9781824

>
>
```

Figure 6-11. Correlation between skipped English subtitles and NRN / RN redundant in R.

```
>
>
> fit1 <- glmer(FC ~ Group + (1|Participant) + (1|AOI), family=poisson(link = 'log'), data=subset(Fix_MW_clean, FC>0))
> fit3 <- glmer(FC ~ Group + Redundant + (1|Participant) + (1|AOI), family=poisson(link = 'log'), data=subset(Fix_MW_clean, FC>0))
> fit5 <- glmer(FC ~ Group*CPS_centred + (1|Participant) + (1|AOI), family=poisson(link = 'log'), data=subset(Fix_MW_clean, FC>0))
> fit6 <- glmer(FC ~ Group*CPS_centred + Redundant + (1|Participant) + (1|AOI),
+               family=poisson(link = 'log'),
+               data=subset(Fix_MW_clean, FC>0),
+               control=glmerControl(optimizer="Nelder_Mead", optCtrl = list(maxfun=1000000)))
> fit7 <- glmer(FC ~ Group*CPS_centred + MW_centered + Redundant + (1|Participant) + (1|AOI),
+               family=poisson(link = 'log'),
+               data=subset(Fix_MW_clean, FC>0),
+               control=glmerControl(optimizer="Nelder_Mead", optCtrl = list(maxfun=1000000)))
> anova(fit1, fit3, fit5, fit6, fit7)
Data: subset(Fix_MW_clean, FC > 0)
Models:
fit1: FC ~ Group + (1 | Participant) + (1 | AOI)
fit5: FC ~ Group * CPS_centred + (1 | Participant) + (1 | AOI)
fit3: FC ~ Group + Redundant + (1 | Participant) + (1 | AOI)
fit6: FC ~ Group * CPS_centred + Redundant + (1 | Participant) + (1 |
fit6:     AOI)
fit7: FC ~ Group * CPS_centred + MW_centered + Redundant + (1 | Participant) +
fit7:     (1 | AOI)
     Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
fit1  4 47706 47735 -23849    47698
fit5  6 47538 47582 -23763    47526  171.37      2     <2e-16 ***
fit3  7 47674 47725 -23830    47660    0.00      1          1
fit6  9 47472 47538 -23727    47454  206.05      2     <2e-16 ***
fit7 10 45728 45801 -22854    45708 1746.22      1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure 6-12. FC model comparison output in R.

```
>
> # add CPS_centred as interaction with Group and MW_centered and Redundant as fixed effect
> fit7 <- glmer(FC ~ Group*CPS_centred + MW_centered + Redundant + (1|Participant) + (1|AOI),
+               family=poisson(link = 'log'),
+               data=subset(Fix_MW_clean, FC>0),
+               control=glmerControl(optimizer="Nelder_Mead", optCtrl = list(maxfun=1000000)))
>
>
> summary(fit7)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: poisson  ( log )
Formula: FC ~ Group * CPS_centred + MW_centered + Redundant + (1 | Participant) +      (1 | AOI)
   Data: subset(Fix_MW_clean, FC > 0)
Control: glmerControl(optimizer = "Nelder_Mead", optCtrl = list(maxfun = 1e+06))

     AIC      BIC   logLik deviance df.resid
 44079.4  44151.7 -22029.7  44059.4    10224

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2405 -0.6830 -0.0902  0.5826  5.0827

Random effects:
 Groups      Name        Variance Std.Dev.
 AOI         (Intercept) 0.02175  0.1475
 Participant (Intercept) 0.07243  0.2691
Number of obs: 10234, groups:  AOI, 403; Participant, 44

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                  1.295666   0.059743  21.687  < 2e-16 ***
GroupEnglish_sub             0.250307   0.082161   3.047  0.00231 **
CPS_centred                 -0.141571   0.010404 -13.607  < 2e-16 ***
MW_centered                  0.456098   0.008476  53.811  < 2e-16 ***
RedundantNRWH               -0.094157   0.023682  -3.976 7.01e-05 ***
RedundantRH                 -0.146810   0.021168  -6.935 4.05e-12 ***
RedundantRN                  0.042795   0.026283   1.628  0.10348
GroupEnglish_sub:CPS_centred -0.008442  0.010234  -0.825  0.40947
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) GrpEn_ CPS_cn MW_cnt RdNRWH RdndRH RdndRN
GrpEnglsh_s -0.691
CPS_centred -0.022  0.006
MW_centered  0.004 -0.018 -0.422
RedndntNRWH -0.125 -0.001  0.017  0.015
RedundantRH -0.150  0.002  0.198 -0.116  0.373
RedundantRN -0.118  0.004  0.075 -0.088  0.307  0.339
GrpEn_:CPS_ -0.002 -0.005 -0.577  0.101  0.020  0.016  0.008
>
```

Figure 6-13. *glmer* output script for in R.

```
>
> mod1 <- glmer(MFD ~ Group + (1|Participant) + (1|AOI),
+              family=Gamma(link = 'log'),
+              data=subset(MFD_clean, FC>0))
> mod2 <- glmer(MFD ~ Group + CPS_centred + Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
> mod3 <- glmer(MFD ~ Group + MW_centered + Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
> mod4 <- glmer(MFD ~ Group + Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
> mod5 <- glmer(MFD ~ Group*Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
> mod6 <- glmer(MFD ~ Group*MW_centered + Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
> mod7 <- glmer(MFD ~ Group + CPS_centred + MW_centered + Redundant + (1|Participant) + (1|AOI),
+                family=Gamma(link ='log'),
+                data=subset(MFD_clean, FC > 0))
>
> #compare models
> anova(mod1, mod2, mod3, mod4, mod5, mod6, mod7)
Data: subset(MFD_clean, FC > 0)
Models:
mod1: MFD ~ Group + (1 | Participant) + (1 | AOI)
mod4: MFD ~ Group + Redundant + (1 | Participant) + (1 | AOI)
mod2: MFD ~ Group + CPS_centred + Redundant + (1 | Participant) + (1 |
mod2:     AOI)
mod3: MFD ~ Group + MW_centered + Redundant + (1 | Participant) + (1 |
mod3:     AOI)
mod6: MFD ~ Group * MW_centered + Redundant + (1 | Participant) + (1 |
mod6:     AOI)
mod7: MFD ~ Group + CPS_centred + MW_centered + Redundant + (1 | Participant) +
mod7:     (1 | AOI)
mod5: MFD ~ Group * Redundant + (1 | Participant) + (1 | AOI)
     Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
mod1  5 119152 119188 -59571   119142
mod4  8 119147 119205 -59565   119131 11.276      3    0.01033 *
mod2  9 119119 119184 -59550   119101 29.812      1  4.760e-08 ***
mod3  9 119127 119192 -59554   119109  0.000      0    1.00000
mod6 10 119081 119154 -59531   119061 47.609      1  5.203e-12 ***
mod7 10 119115 119187 -59547   119095  0.000      0    1.00000
mod5 11 119149 119228 -59563   119127  0.000      1    1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Figure 6-16. MFD model comparison output in R.

```
>
> mod6 <- glmer(MFD ~ Group*MW_centered + Redundant + (1|Participant) + (1|AOI),
+               family=Gamma(link ='log'),
+               data=subset(MFD_clean, FC > 0))
> summary(mod6)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Gamma  ( log )
Formula: MFD ~ Group * MW_centered + Redundant + (1 | Participant) + (1 |      AOI)
   Data: subset(MFD_clean, FC > 0)

     AIC      BIC   logLik deviance df.resid
119081.3 119153.6 -59530.7 119061.3    10165

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7143 -0.6530 -0.1078  0.5280  6.1847

Random effects:
 Groups      Name        Variance Std.Dev.
 AOI         (Intercept) 0.004028 0.06347
 Participant (Intercept) 0.005757 0.07587
 Residual                0.090086 0.30014
Number of obs: 10175, groups:  AOI, 403; Participant, 44

Fixed effects:
                              Estimate Std. Error t value Pr(>|z|)
(Intercept)                   5.581660   0.042327 131.869  < 2e-16 ***
GroupEnglish_sub              0.138193   0.057157   2.418  0.01562 *
MW_centered                  -0.048444   0.006067  -7.985 1.41e-15 ***
RedundantNRWH                -0.049643   0.015264  -3.252  0.00115 **
RedundantRH                  -0.008912   0.013552  -0.658  0.51076
RedundantRN                  -0.015423   0.017837  -0.865  0.38722
GroupEnglish_sub:MW_centered  0.045385   0.006565   6.913 4.75e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) GrpEn_ MW_cnt RdNRWH RdndRH RdndRN
GrpEnglsh_s -0.686
MW_centered  0.027  0.001
RedndntNRWH -0.115 -0.003 -0.015
RedundantRH -0.142  0.000 -0.045  0.393
RedundantRN -0.103 -0.001 -0.088  0.296  0.316
GrpEng_:MW_  0.001 -0.021 -0.609  0.011  0.016  0.041
>
```

Figure 6-17. MFD output computation for mod6 in R.

```
>
> # order of coefficients
> fixef(mod6)
                (Intercept)            GroupEnglish_sub                   MW_centered               RedundantNRWH
                5.581660169                 0.138192767                  -0.048444067                -0.049642936
                RedundantRH                 RedundantRN GroupEnglish_sub:MW_centered
               -0.008912333                -0.015422734                   0.045385326
>
> #Intercept is 265.51ms, MFD for reading Chinese subtitles is 265.51ms with average MW for Chinese.
> exp(eff6[1])
(Intercept)
    265.512
>
> ## MFD in reading English subtitles with average MW for English is 304.86ms,
> ## significantly longer MFD than reading Chinese subtitles with average MW for Chinese.
> exp(eff6[1] + eff6[2])
(Intercept)
   304.8601
>
>
> ## MFD is 252.63ms in reading Chinese subtitles (average MW for Chinese) with NRWH redundant.
> ## significantly shorter MFD with the presence of hand movement with no redunant.
> exp(eff6[1] + eff6[4])
(Intercept)
   252.6531
>
>
> ## MFD in reading Chinese subtitles with reduced MW is 252.97ms,
> ## significantly shorter MFD when mean word length is reduced in Chinese subtitles.
> exp(eff6[1] + eff6[3])
(Intercept)
   252.9561
>
>
> ## MFD is 303.92ms in reading English subtitles with increase MW
> ## significantly longer MFD when mean word length is increase in English subtitles
> exp(eff6[1] + eff6[2] + eff6[3] + eff6[7])
(Intercept)
    303.929
>
```

Figure 6-18. MFD mod6 detail computation outcome script in R.

```
>
> qqnorm(resid(mod6))
> qqline(resid(mod6), col='red')
> plot(fitted.values(mod6), resid(mod6))
> abline(0,0, col=2)
> hist(resid(mod6))
> cor(resid(mod6), fitted.values(mod6))
[1] 0.06895747
>
```

Figure 6-22. MFD residual script in R.

```
> anova(fit1, fit2, fit3, fit4, fit5, fit6, fit7, fit8, fit9, fit10, fit11, fit12)
Data: subset(DTpercent_clean, DTpercent > 0)
Models:
fit1: DTpercent ~ Group + (1 | Participant) + (1 | AOI), zi=~Group, disp=~1
fit6: DTpercent ~ Group + CPS_centred + (1 | Participant) + (1 | AOI), zi=~Group + CPS_centred, disp=~1
fit12: DTpercent ~ Group + MW_centered + (1 | Participant) + (1 | AOI), zi=~Group + MW_centered, disp=~1
fit4: DTpercent ~ Group * CPS_centred + (1 | Participant) + (1 | AOI), zi=~Group * CPS_centred, disp=~1
fit8: DTpercent ~ Group + MW_centered + CPS_centred + (1 | Participant) + , zi=~Group + MW_centered + CPS_centred, disp=~1
fit8:     (1 | AOI), zi=~Group * MW_centered, disp=~1
fit9: DTpercent ~ Group * MW_centered + (1 | Participant) + (1 | AOI), zi=~Group + Redundant, disp=~1
fit2: DTpercent ~ Group + Redundant + (1 | Participant) + (1 | AOI), zi=~Group + MW_centered + Redundant, disp=~1
fit7: DTpercent ~ Group + MW_centered + Redundant + (1 | Participant) + , zi=~Group * CPS_centred + Redundant, disp=~1
fit7:     (1 | AOI), zi=~Group * MW_centered + Redundant, disp=~1
fit5: DTpercent ~ Group * CPS_centred + Redundant + (1 | Participant) + , zi=~Group + CPS_centred + MW_centered + Redundant, disp=~1
fit5:     (1 | AOI), zi=~Group * Redundant, disp=~1
fit10: DTpercent ~ Group * MW_centered + Redundant + (1 | Participant) + , zi=~Group, disp=~1
fit10:     (1 | AOI), zi=~Group + CPS_centred, disp=~1
fit11: DTpercent ~ Group + CPS_centred + MW_centered + Redundant + (1 | , zi=~Group + MW_centered, disp=~1
fit11:     Participant) + (1 | AOI), zi=~Group * CPS_centred, disp=~1
fit3: DTpercent ~ Group * Redundant + (1 | Participant) + (1 | AOI), zi=~Group + MW_centered + CPS_centred, disp=~1
      Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
fit1   7 88646 88697 -44316    88632
fit6   9 88401 88466 -44191    88383 249.155      2  < 2.2e-16 ***
fit12  9 88470 88535 -44226    88452   0.000      0          1
fit4  11 88380 88460 -44179    88358  94.125      2  < 2.2e-16 ***
fit8  11 88352 88431 -44165    88330  28.677      0  < 2.2e-16 ***
fit9  11 88474 88553 -44226    88452   0.000      0          1
fit2  13 88540 88634 -44257    88514   0.000      2          1
fit7  15 88327 88436 -44149    88297 216.470      2  < 2.2e-16 ***
fit5  17 88297 88420 -44132    88263  33.763      2  4.661e-08 ***
fit10 17 88330 88453 -44148    88296   0.000      0          1
fit11 17 88238 88360 -44102    88204  92.818      0  < 2.2e-16 ***
fit3  19 88543 88680 -44253    88505   0.000      2          1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure 6-24. Output script for comparing 12 potential models in calculating DT%.

```
> # add MW, CPS and redundant as fixed effect
> fit11 <- glmmTMB(DTpercent ~ Group + CPS_centred + MW_centered + Redundant + (1|Participant) + (1|AOI),
+                  ziformula = ~ Group + CPS_centred + MW_centered + Redundant,
+                  family=Gamma(link ='log'),
+                  data=subset(DTpercent_clean, DTpercent > 0))
> summary(fit11)
 Family: Gamma  ( log )
Formula:          DTpercent ~ Group + CPS_centred + MW_centered + Redundant + (1 |      Participant) + (1 | AOI)
Zero inflation:             ~Group + CPS_centred + MW_centered + Redundant
Data: subset(DTpercent_clean, DTpercent > 0)

     AIC      BIC   logLik deviance df.resid
 88237.5  88360.4 -44101.8  88203.5    10158

Random effects:

Conditional model:
 Groups      Name        Variance Std.Dev.
 Participant (Intercept) 0.08603  0.2933
 AOI         (Intercept) 0.02915  0.1707
Number of obs: 10175, groups:  Participant, 44; AOI, 403

Dispersion estimate for Gamma family (sigma^2): 0.259

Conditional model:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      3.496253   0.064913   53.86  < 2e-16 ***
GroupEnglish_sub 0.360304   0.089434    4.03 5.61e-05 ***
CPS_centred      0.093502   0.009264   10.09  < 2e-16 ***
MW_centered      0.085916   0.009210    9.33  < 2e-16 ***
RedundantNRWH   -0.186789   0.025837   -7.23 4.85e-13 ***
RedundantRH     -0.253150   0.023498  -10.77  < 2e-16 ***
RedundantRN     -0.020584   0.029978   -0.69    0.492
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Zero-inflation model:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -2.099e+01  7.444e+02  -0.028    0.978
GroupEnglish_sub -8.496e-01  1.104e+03  -0.001    0.999
CPS_centred      -2.336e-02  5.442e+02   0.000    1.000
MW_centered      -1.976e-03  5.458e+02   0.000    1.000
RedundantNRWH    -5.316e-01  1.461e+03   0.000    1.000
RedundantRH      -7.005e-01  1.305e+03  -0.001    1.000
RedundantRN      -4.222e-01  1.615e+03   0.000    1.000
>
```

Figure 6-25. Detail computation for fit11 using *glmmTMB* in calculating DT%.

```
>
> fit1 <- glmer(TTFF ~ Group + (1|Participant) + (1|video/AOI),
+               family=Gamma(link = 'log'),
+               data=subset(TTFF_clean, TTFF>0))
> fit2 <- glmer(TTFF ~ Group + Redundant + (1|Participant) + (1|video/AOI),
+               family=Gamma(link = 'log'),
+               data=subset(TTFF_clean, TTFF>0))
> fit3 <- glmer(TTFF ~ Group*Redundant + (1|Participant) + (1|video/AOI),
+               family=Gamma(link = 'log'),
+               data=subset(TTFF_clean, TTFF>0))
>
>
> anova(fit1, fit2, fit3)
Data: subset(TTFF_clean, TTFF > 0)
Models:
fit1: TTFF ~ Group + (1 | Participant) + (1 | video/AOI)
fit2: TTFF ~ Group + Redundant + (1 | Participant) + (1 | video/AOI)
fit3: TTFF ~ Group * Redundant + (1 | Participant) + (1 | video/AOI)
     Df    AIC    BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
fit1  6 157869 157913 -78928   157857
fit2  9 157823 157888 -78902   157805 52.2945      3  2.592e-11 ***
fit3 12 157824 157911 -78900   157800  4.9764      3     0.1735
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
>
```

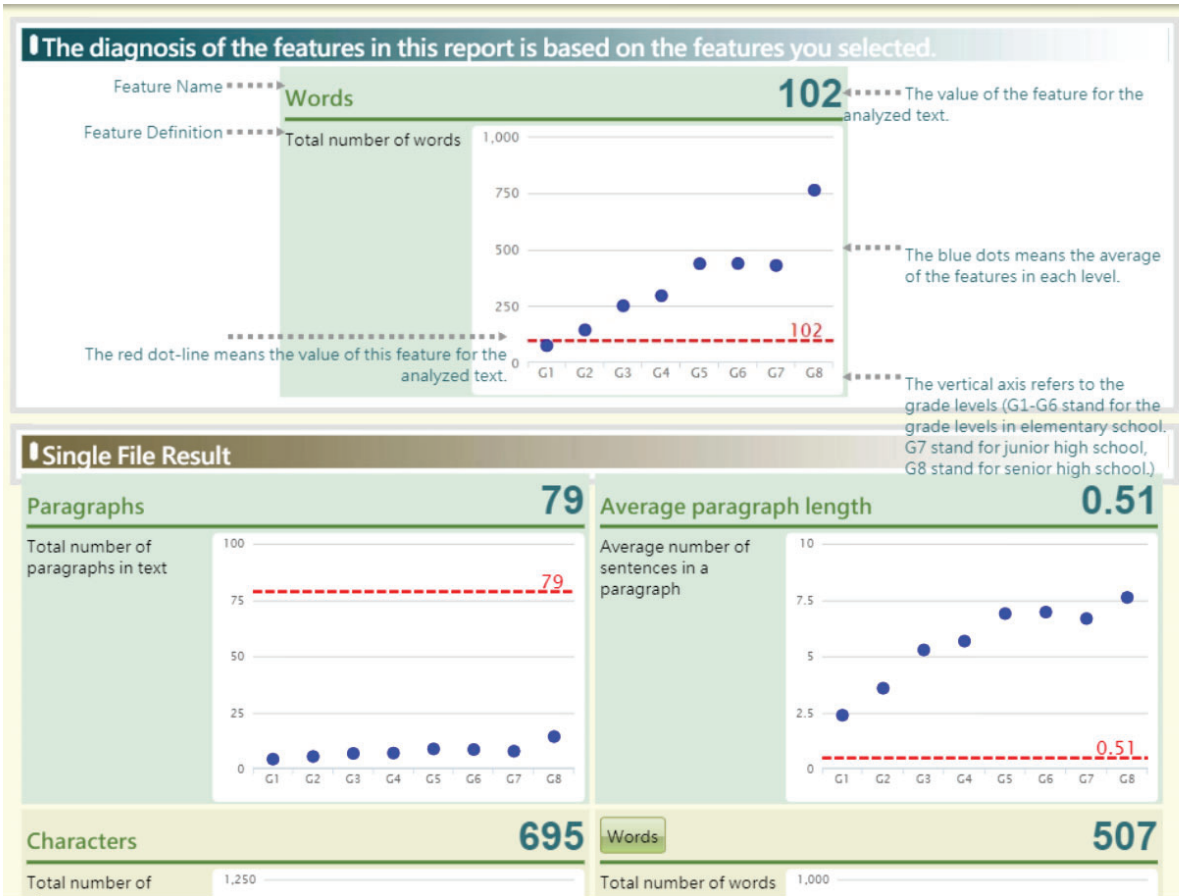
Figure 6-32. Model comparison for TTFF in R.

```
> fit2 <- glmer(TTFF ~ Group + Redundant + (1|Participant) + (1|video/AOI),
+               family=Gamma(link = 'log'),
+               data=subset(TTFF_clean, TTFF>0))
> summary(fit2)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: Gamma  ( log )
Formula: TTFF ~ Group + Redundant + (1 | Participant) + (1 | video/AOI)
   Data: subset(TTFF_clean, TTFF > 0)

     AIC      BIC   logLik deviance df.resid
157822.6 157888.1 -78902.3 157804.6    10743

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.0282 -0.6387 -0.3343  0.2842  9.0873

Random effects:
 Groups      Name        Variance Std.Dev.
 AOI:video   (Intercept) 0.076284 0.27620
 Participant (Intercept) 0.132867 0.36451
 video       (Intercept) 0.007128 0.08443
 Residual                0.945390 0.97231
Number of obs: 10752, groups:  AOI:video, 403; Participant, 44; video, 5

Fixed effects:
                 Estimate Std. Error t value Pr(>|z|)
(Intercept)       6.37010    0.09339  68.207  < 2e-16 ***
GroupEnglish_sub -0.14792    0.11495  -1.287 0.198141
RedundantNRWH     0.15523    0.04228   3.671 0.000241 ***
RedundantRH       0.27327    0.03774   7.240 4.48e-13 ***
RedundantRN       0.07396    0.04997   1.480 0.138907
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) GrpEn_ RdNRWH RdndRH
GrpEnglsh_s -0.615
RedndntNRWH -0.145 -0.002
RedundantRH -0.180  0.000  0.389
RedundantRN -0.141  0.002  0.292  0.346
>
```

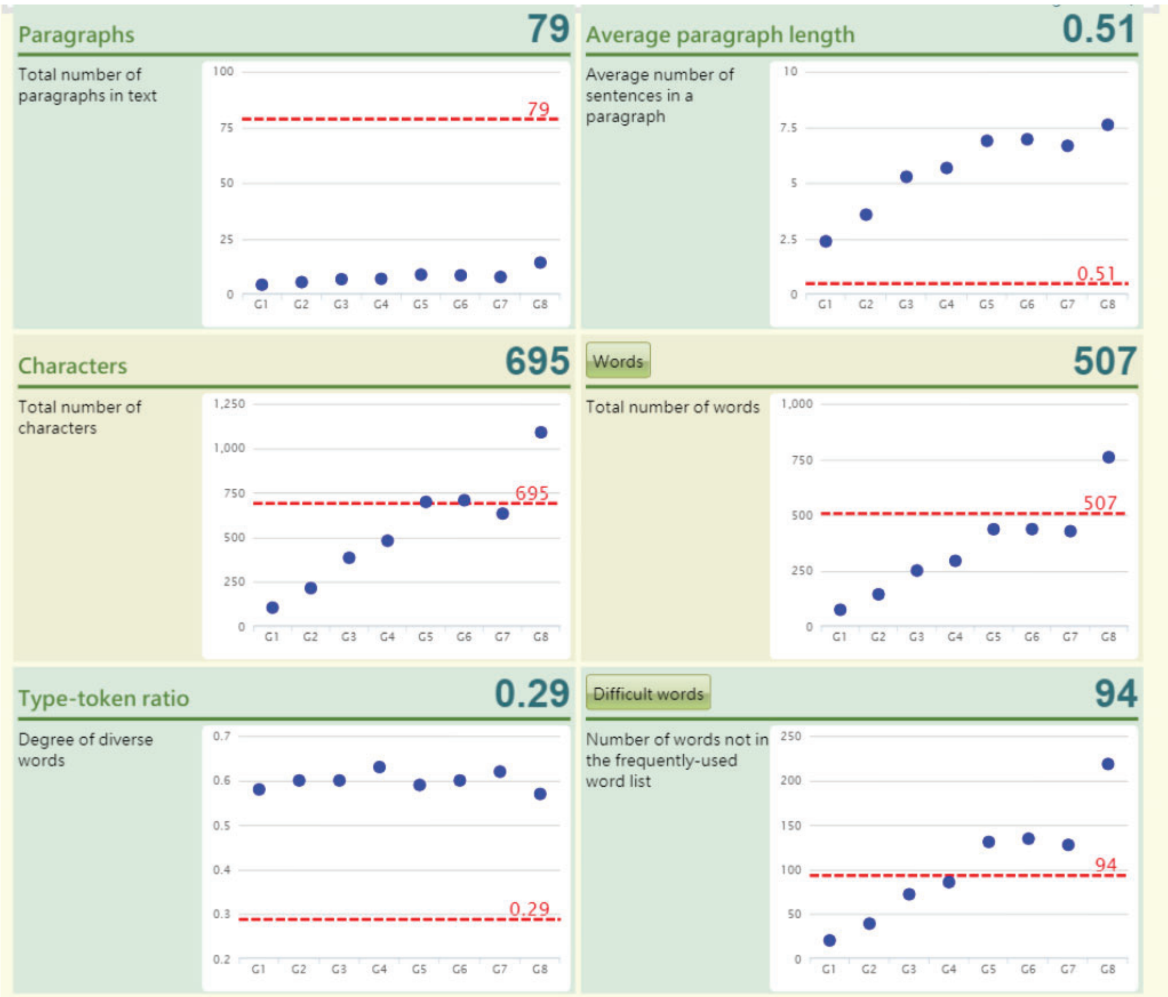
Figure 6-33. *glmer* output for TTFF analysis in R.

Appendix E

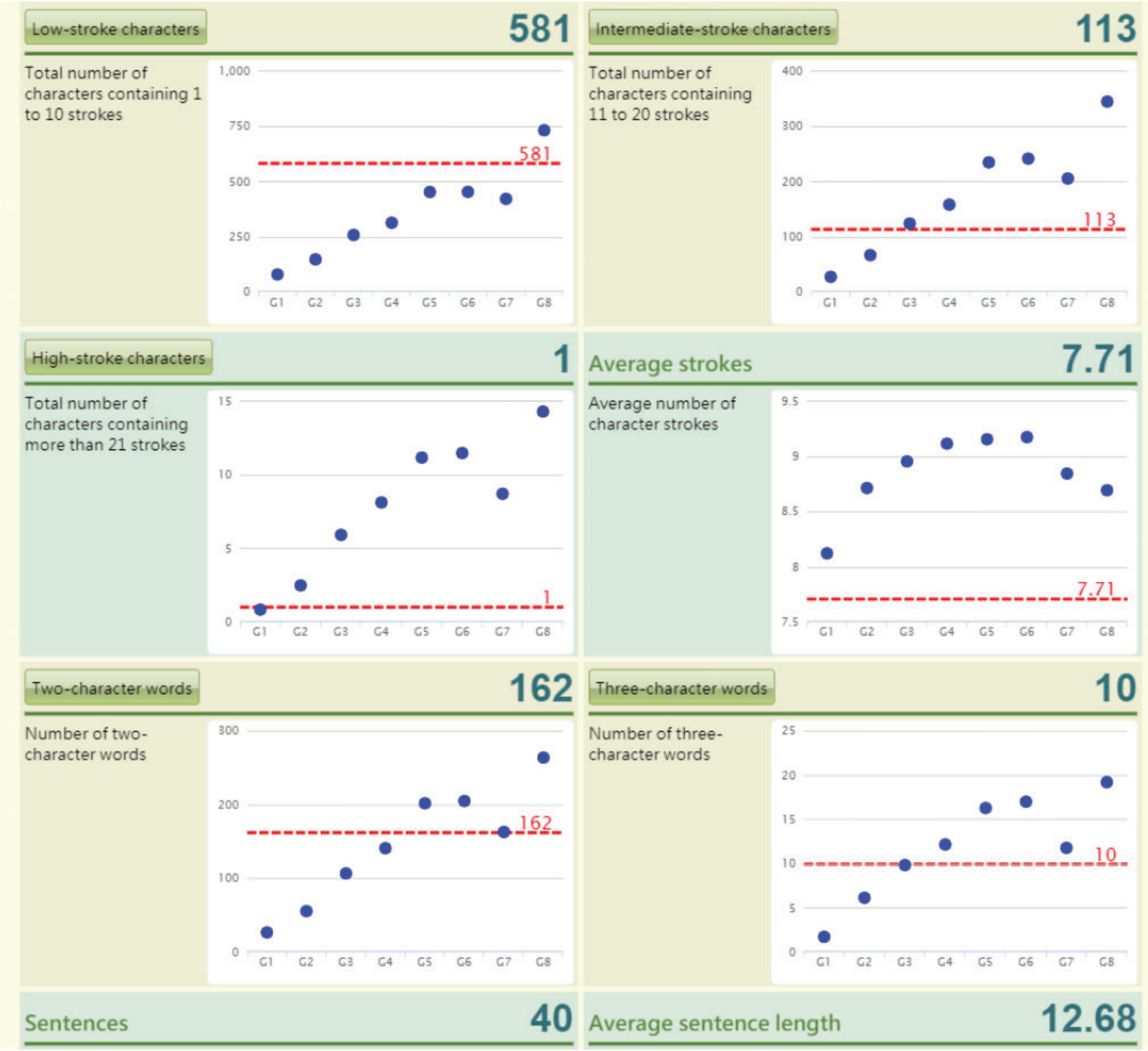
<u>Video 1 text analysis</u>



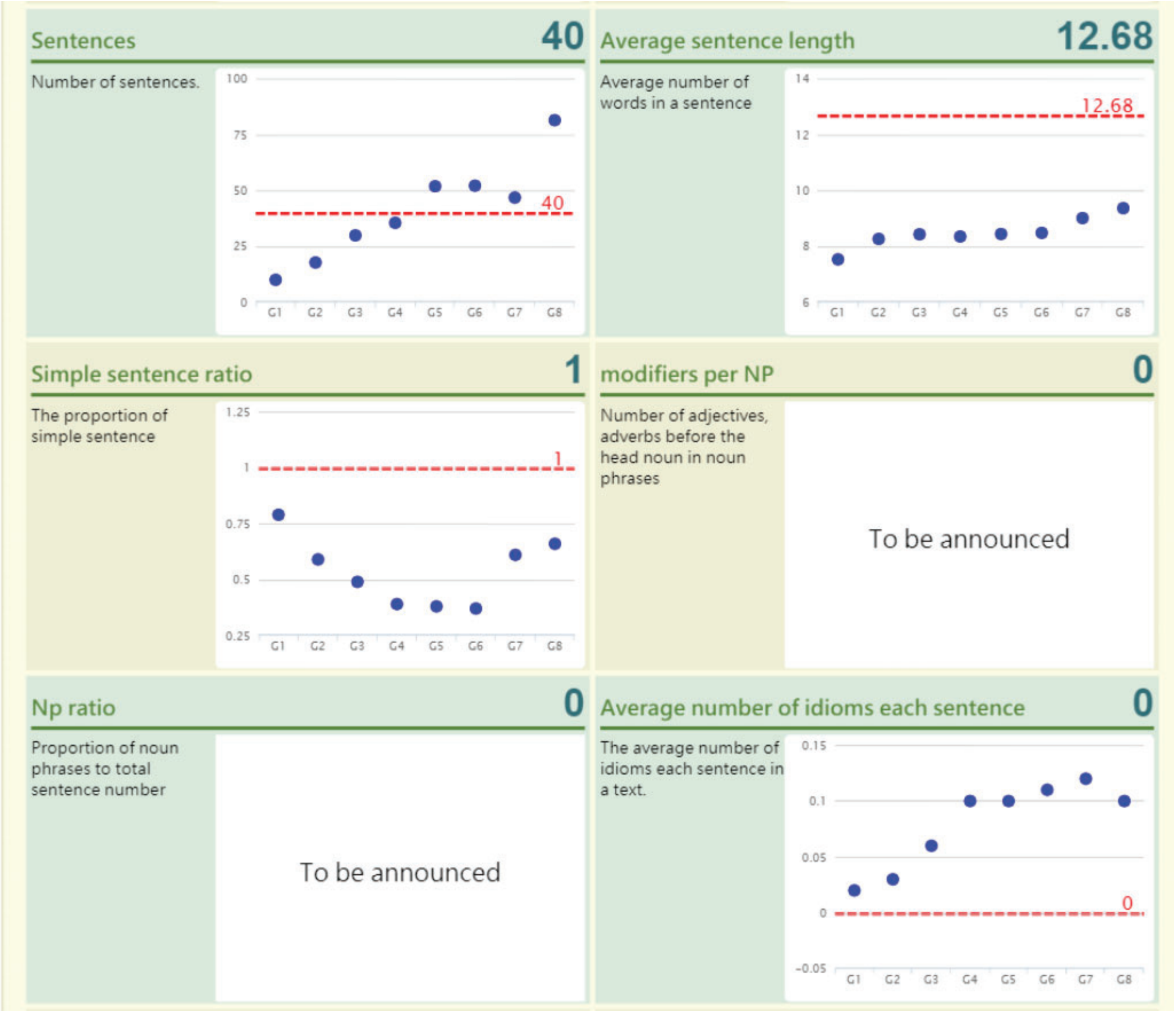
Video 1 text analysis v1.1



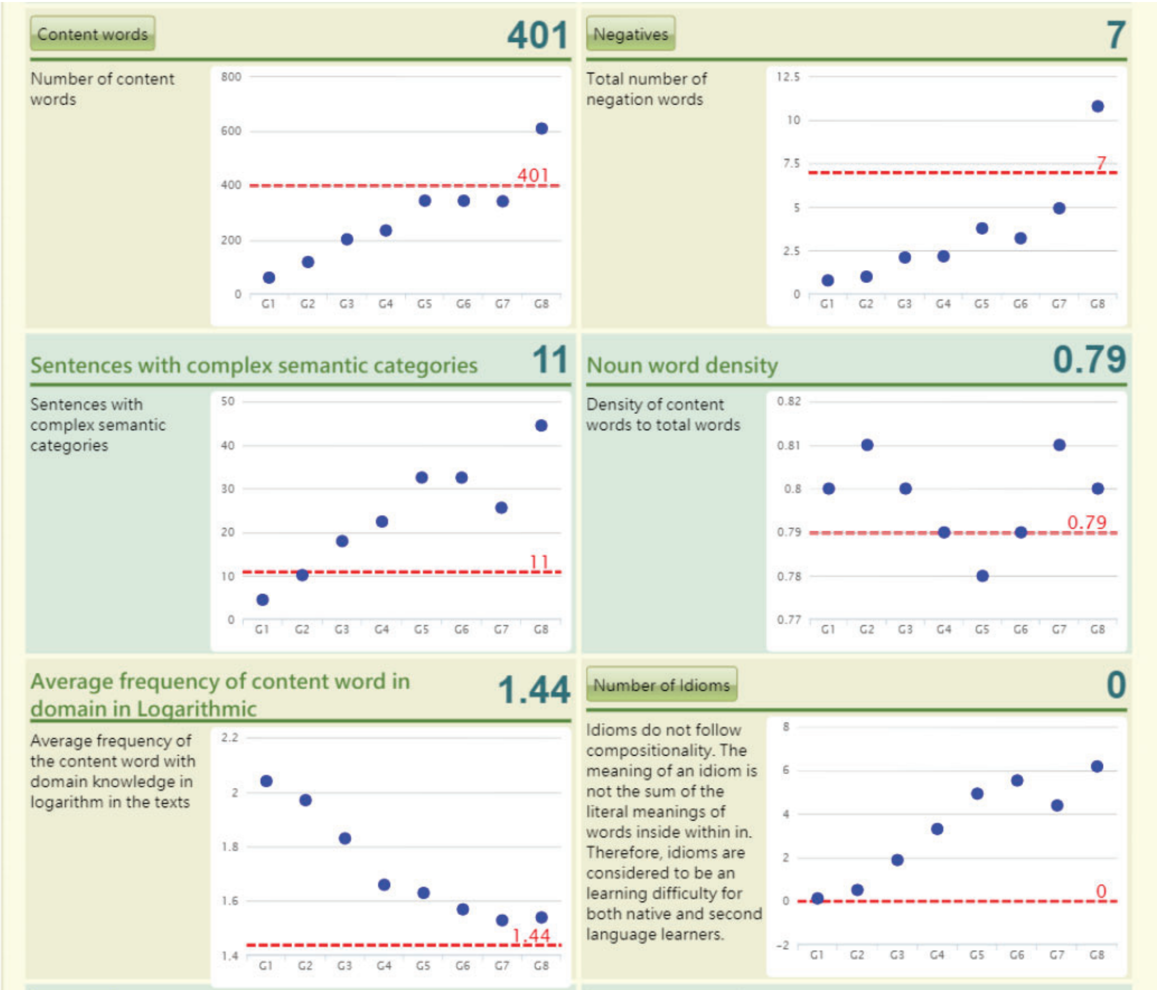
Video 1 text analysis v1.2



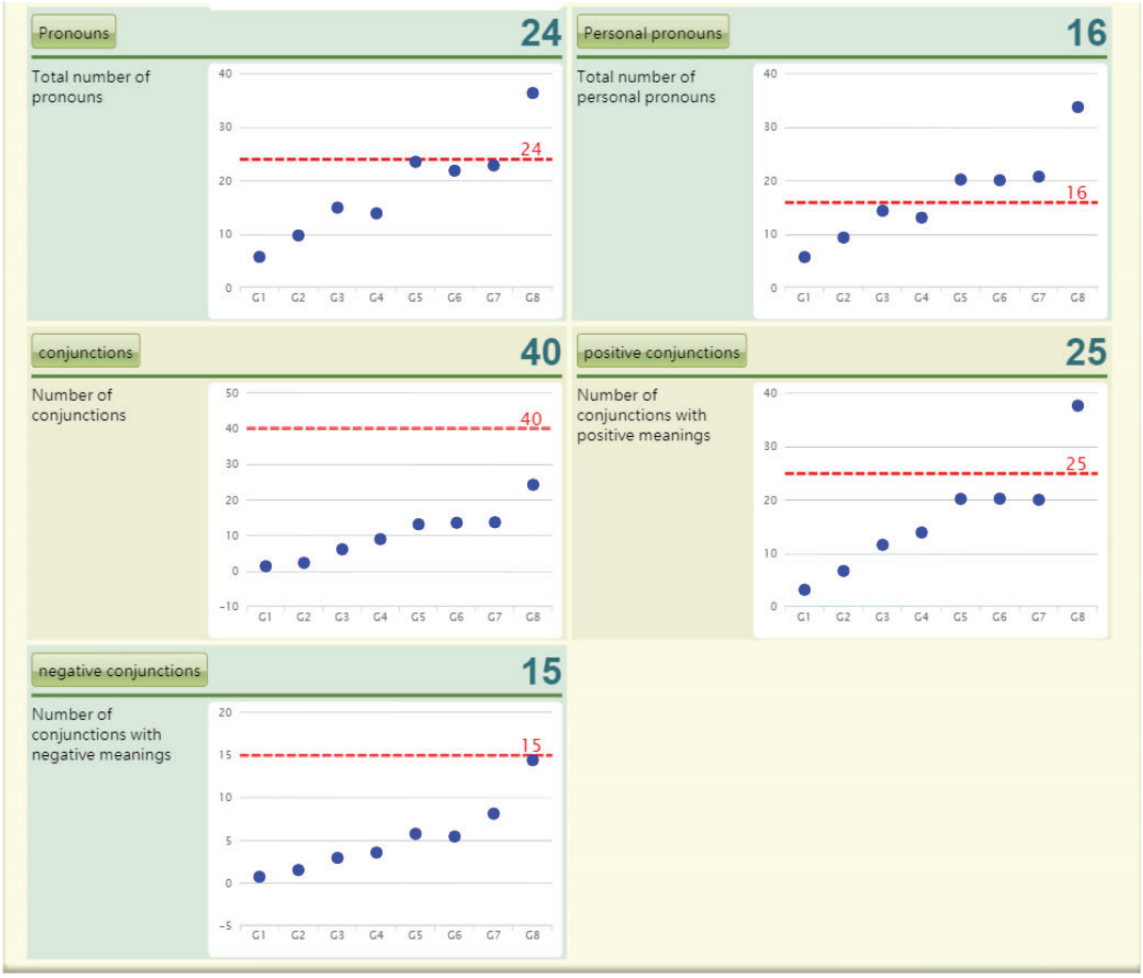
Video 1 text analysis v1.3



Video 1 text analysis v1.4

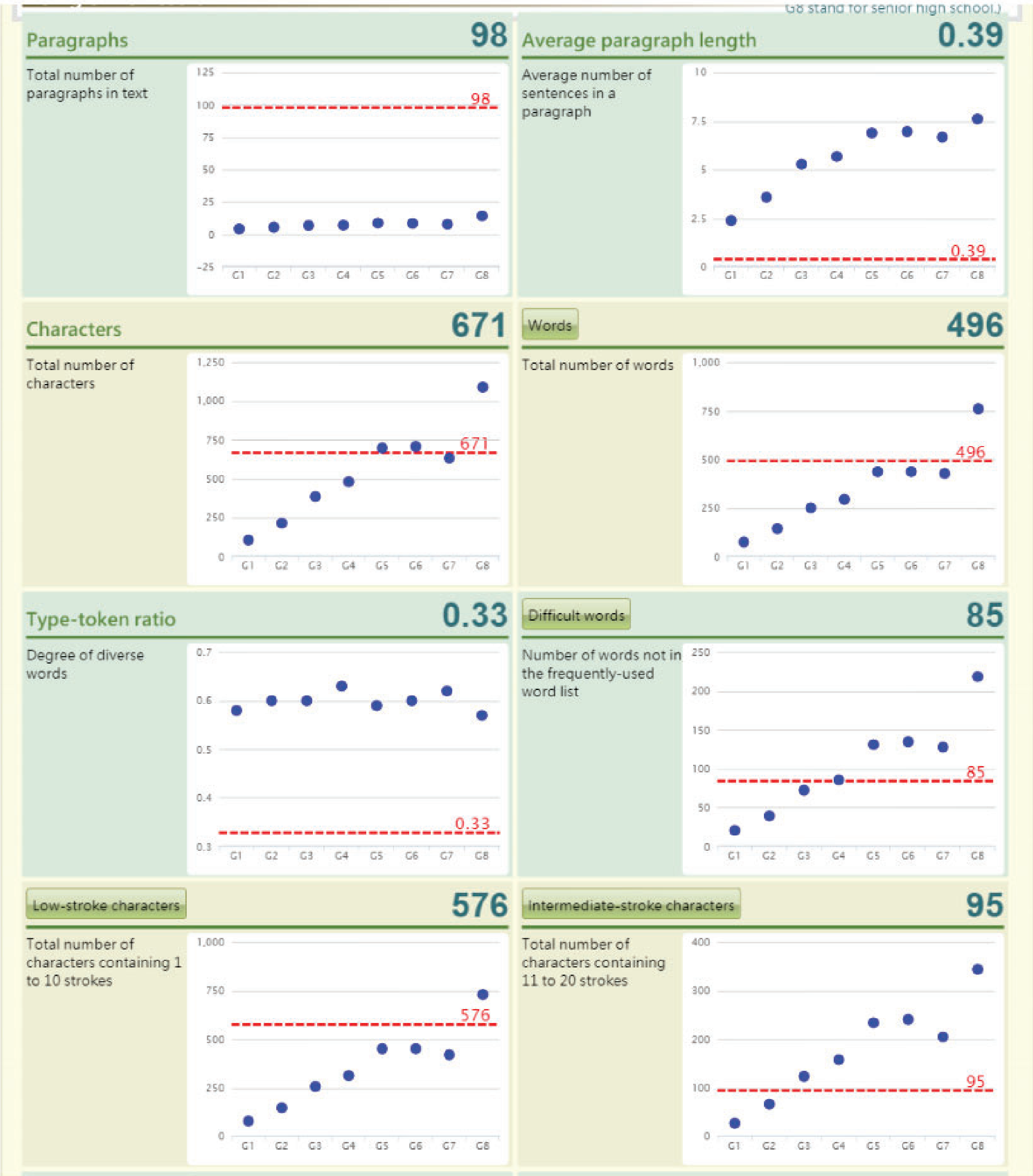


Video 1 text analysis v1.5

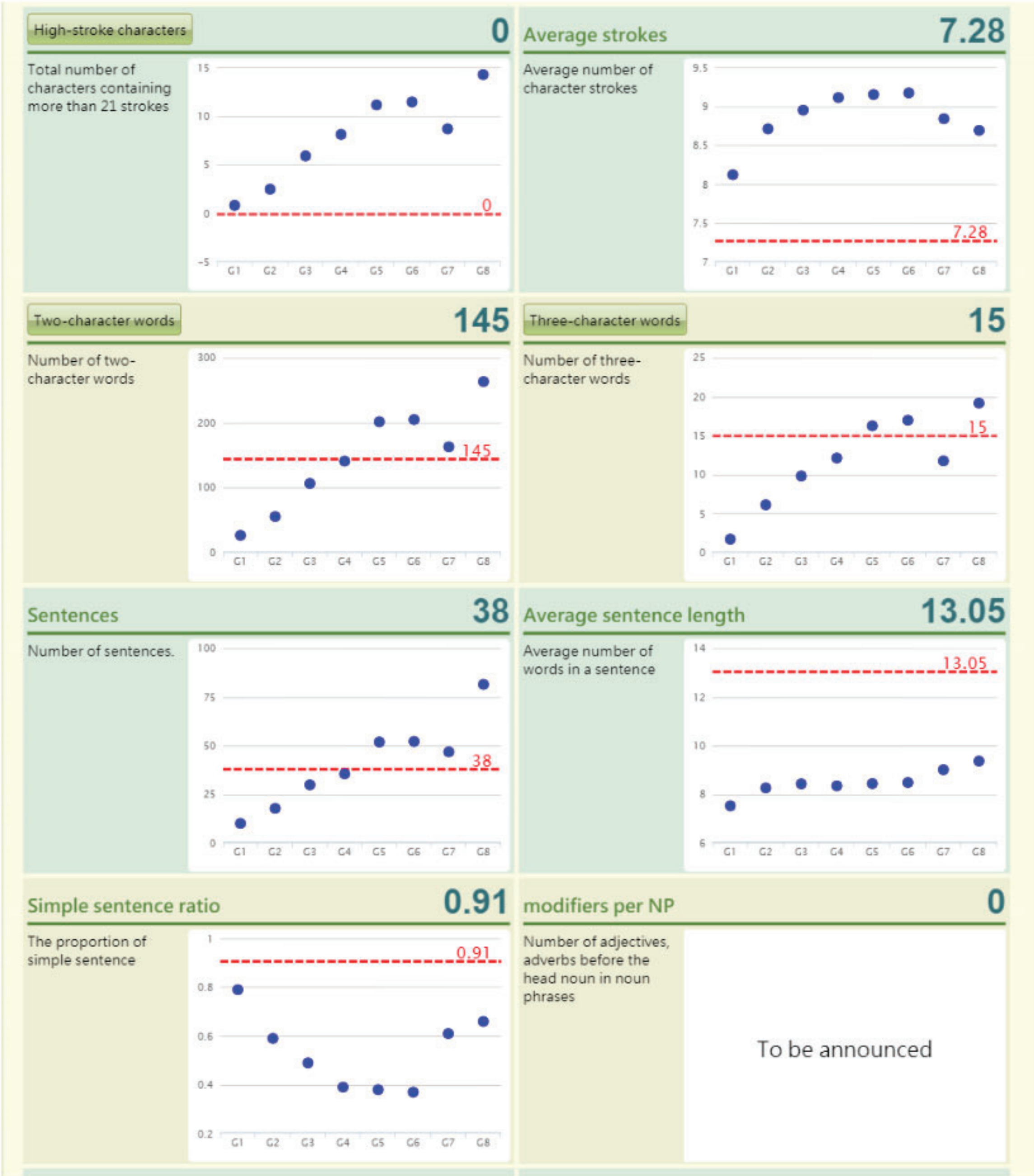


Video 1 text analysis v1.6

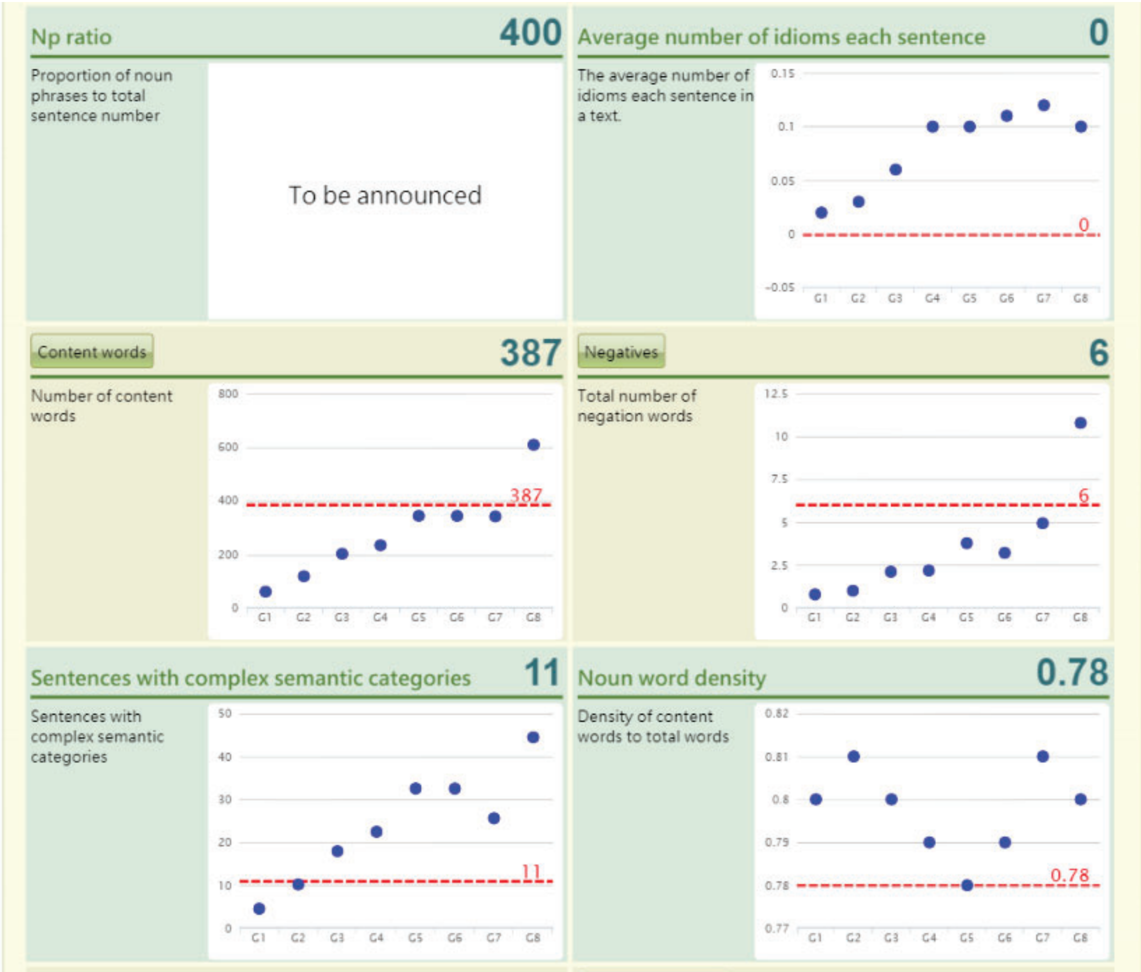
Video 2 text analysis



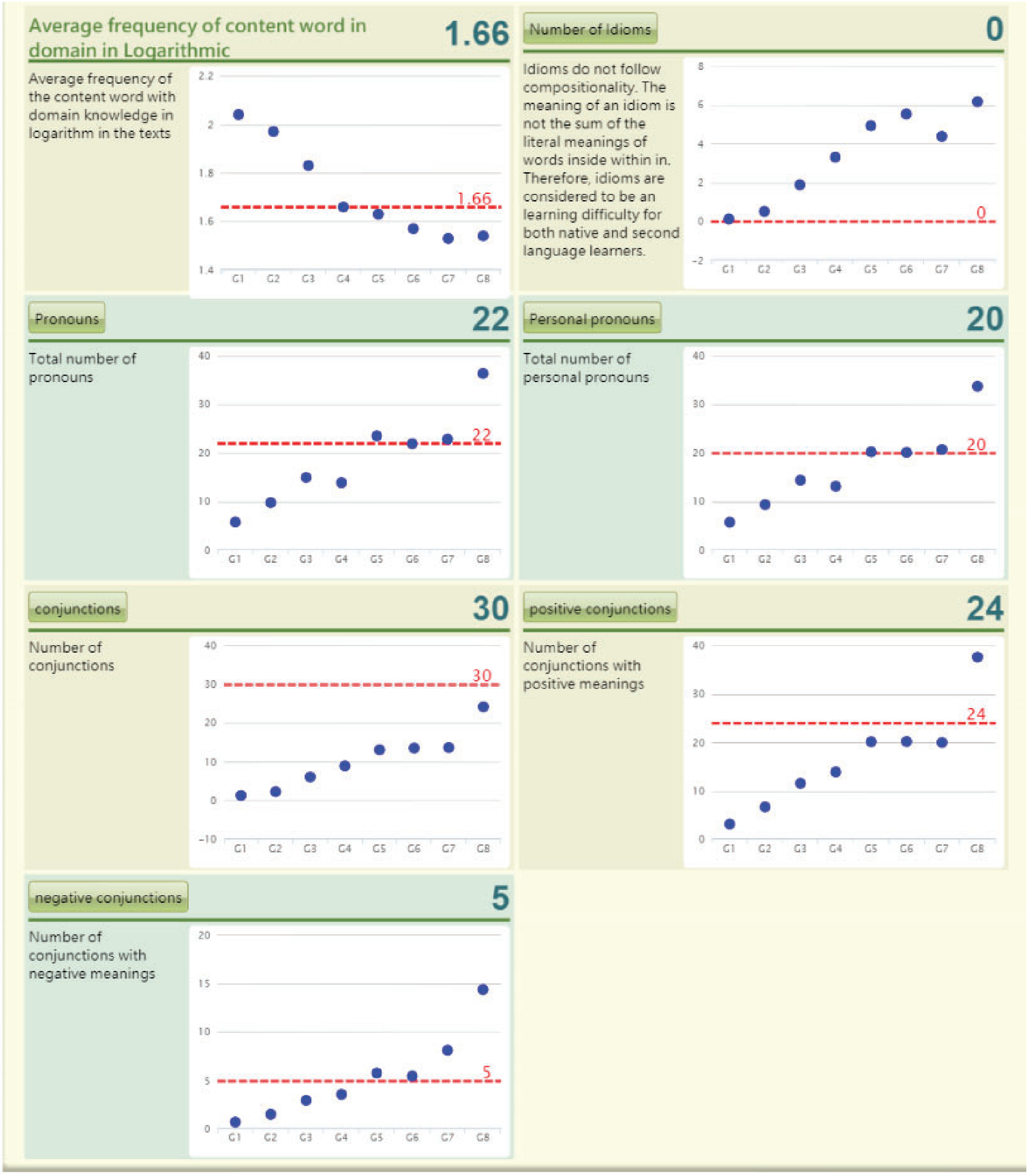
Video 2 text analysis v2.1



Video 2 text analysis v2.2

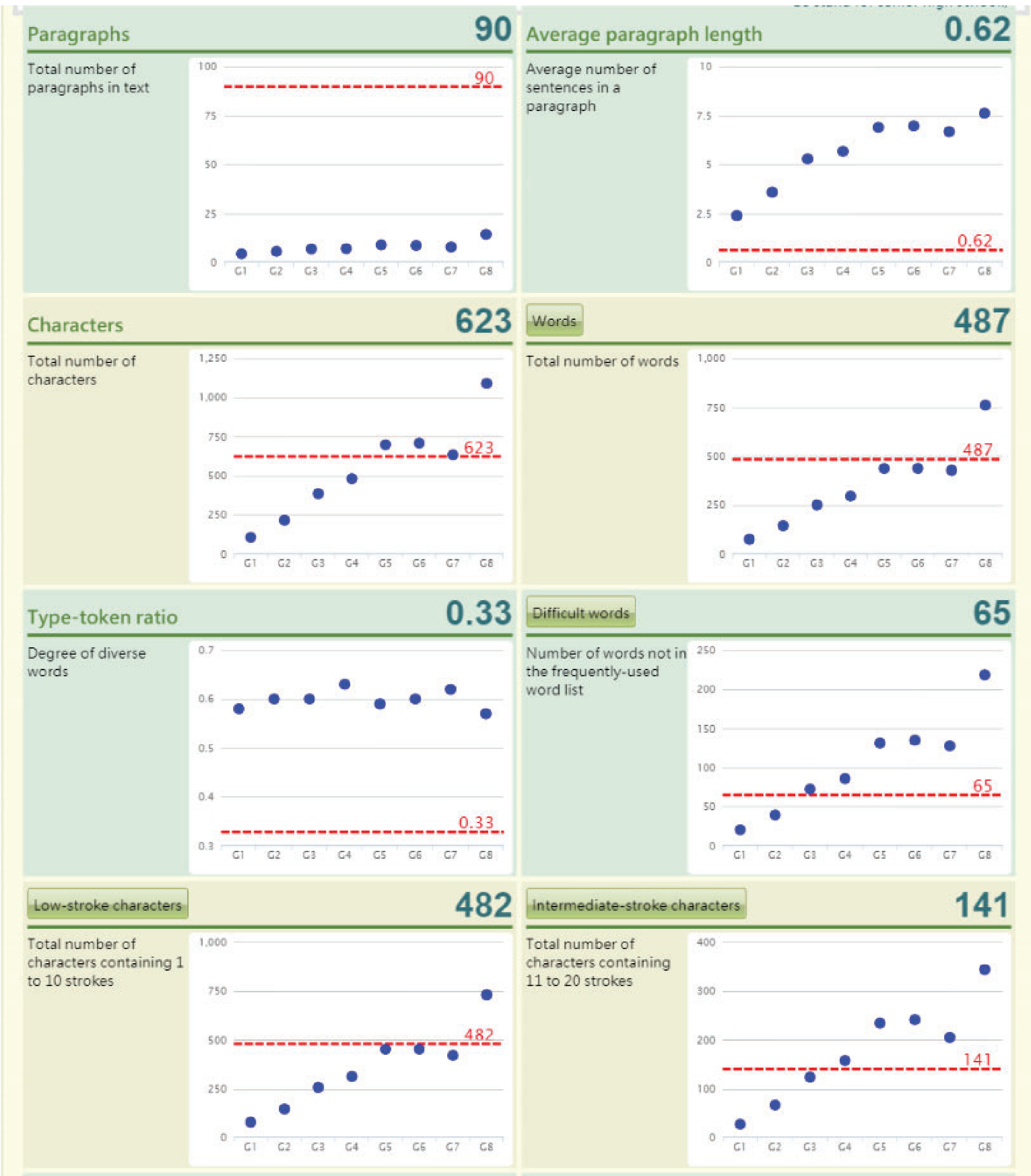


Video 2 text analysis v2.3

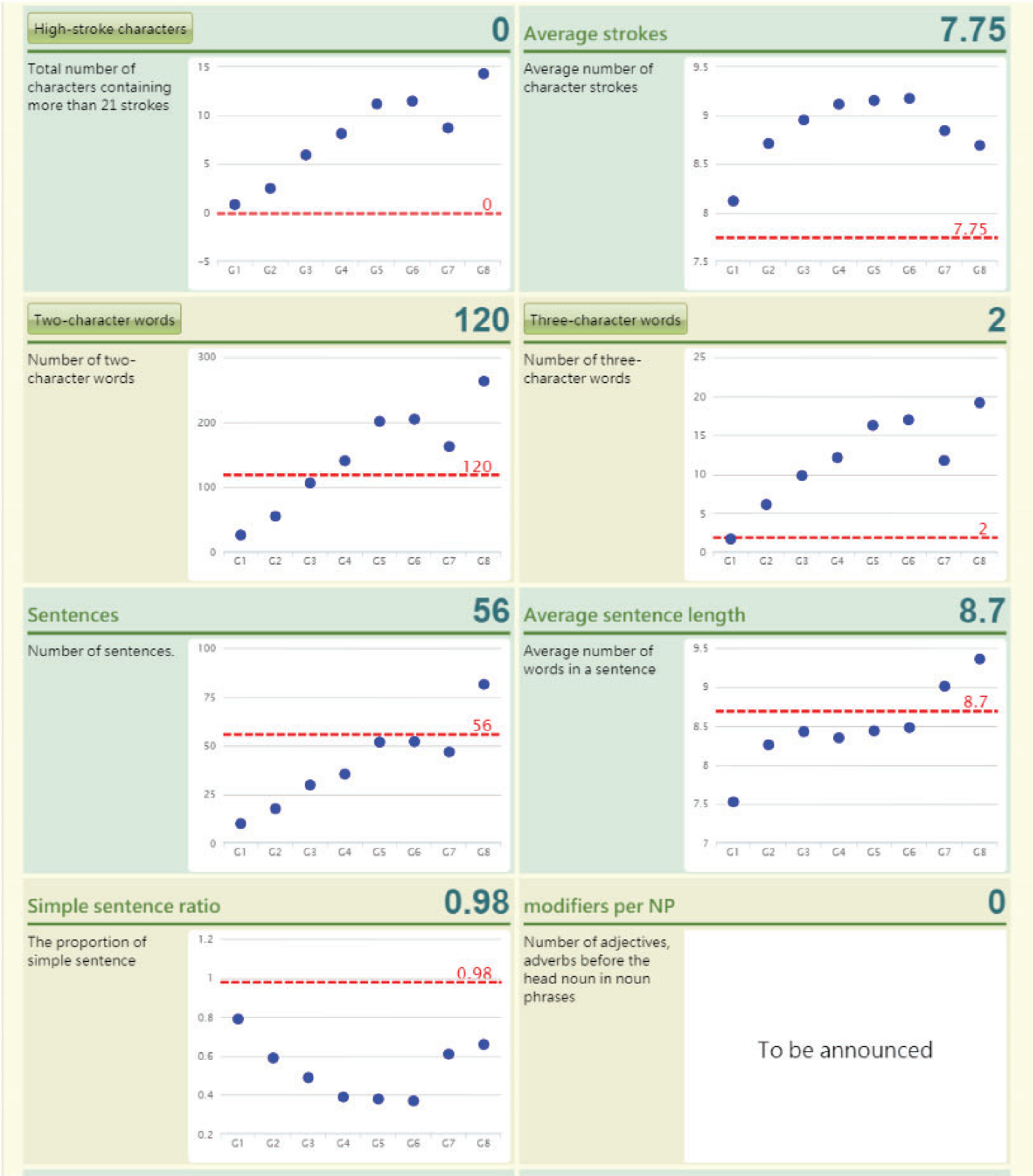


Video 2 text analysis v2.4

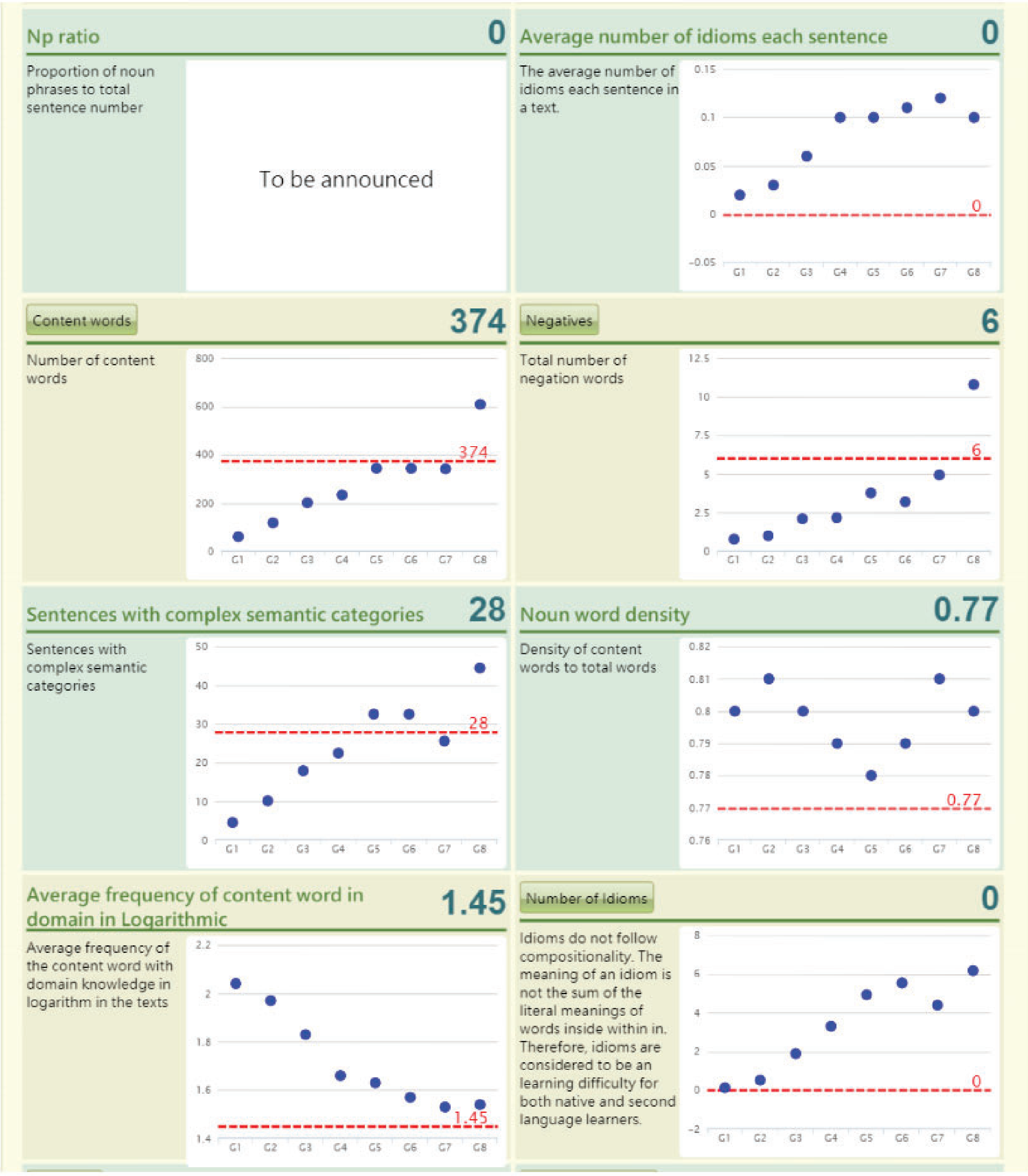
Video 3 text analysis



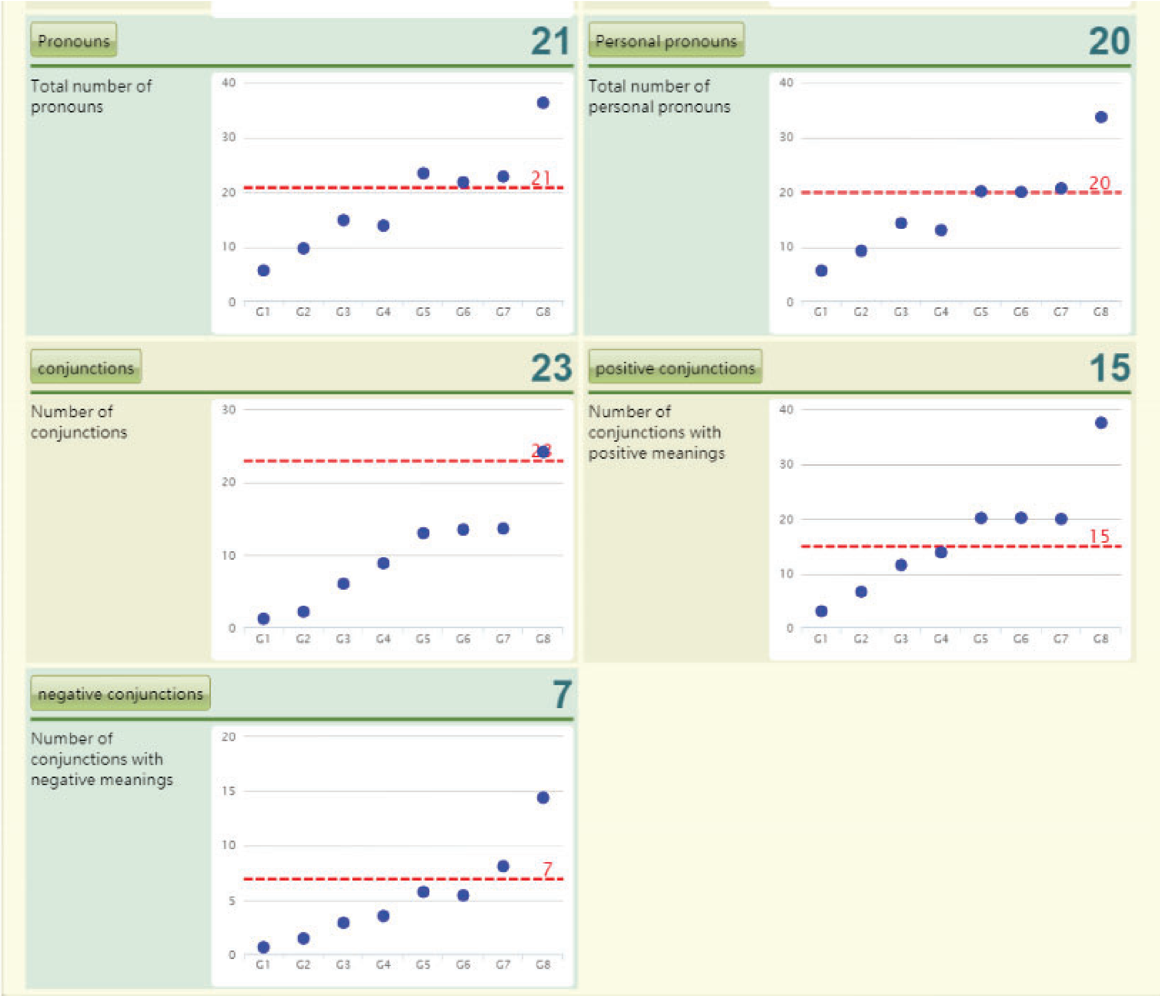
Video 3 text analysis v3.1



Video 3 text analysis v3.2

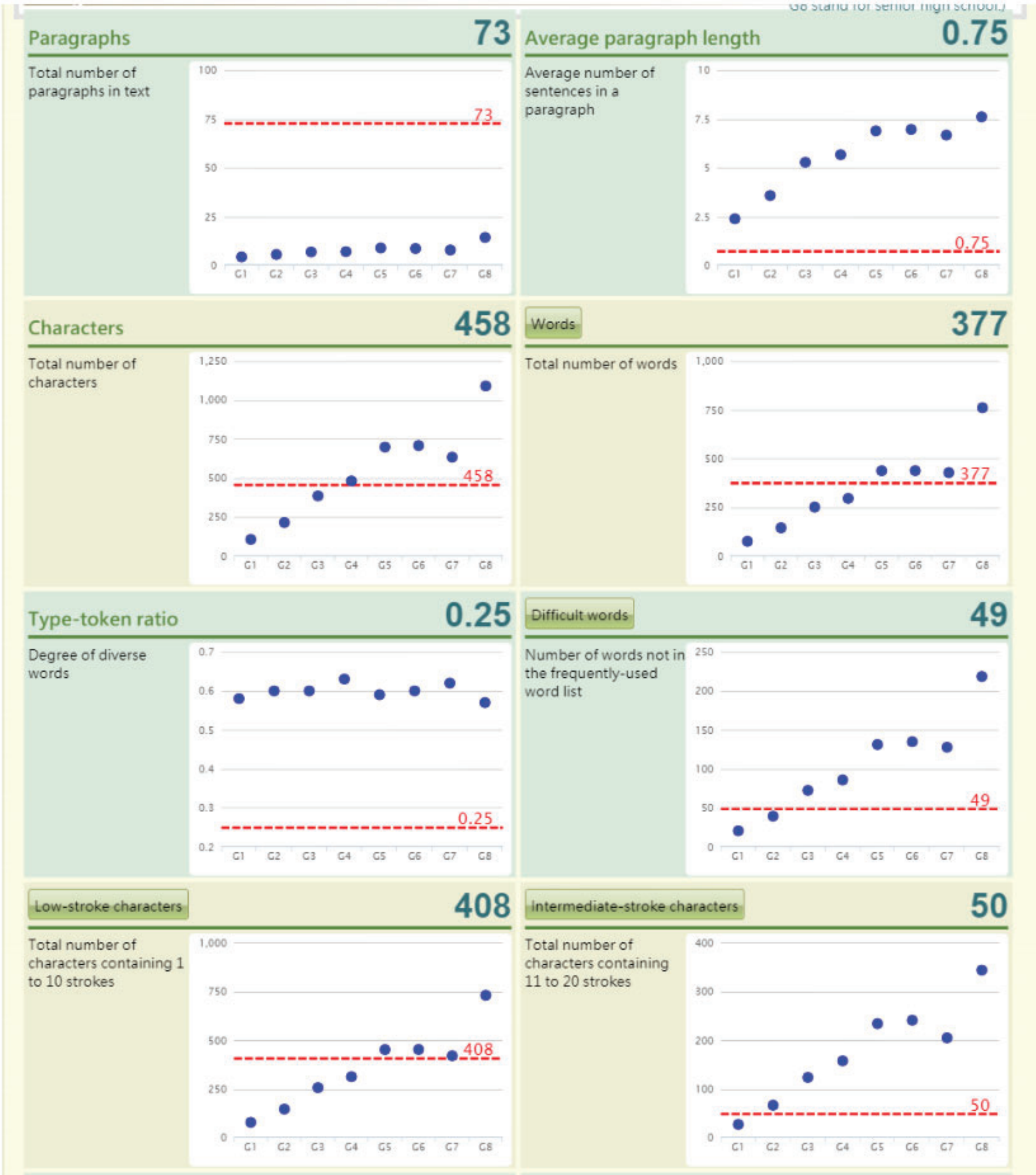


Video 3 text analysis v3.3

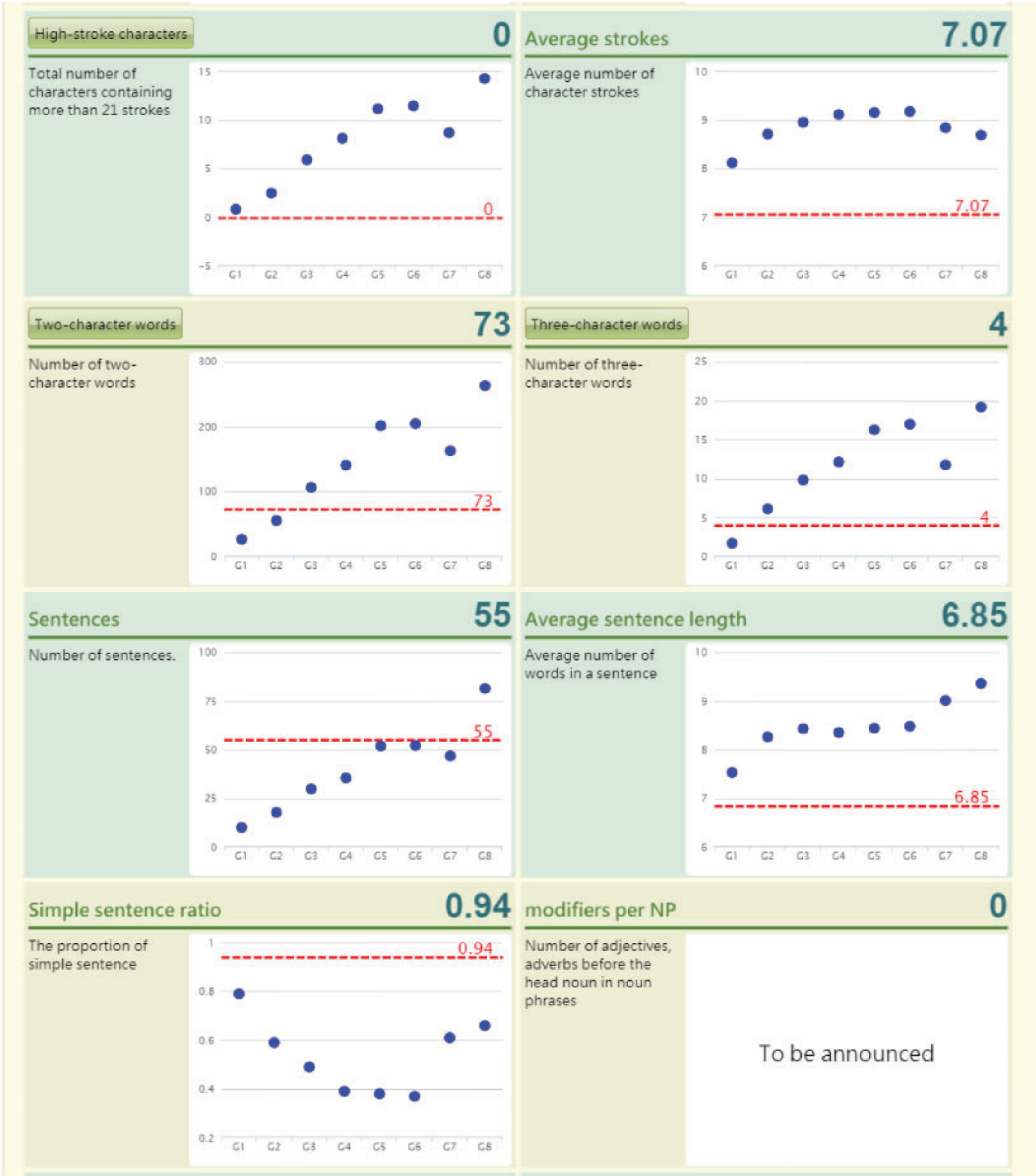


Video 3 text analysis v3.4

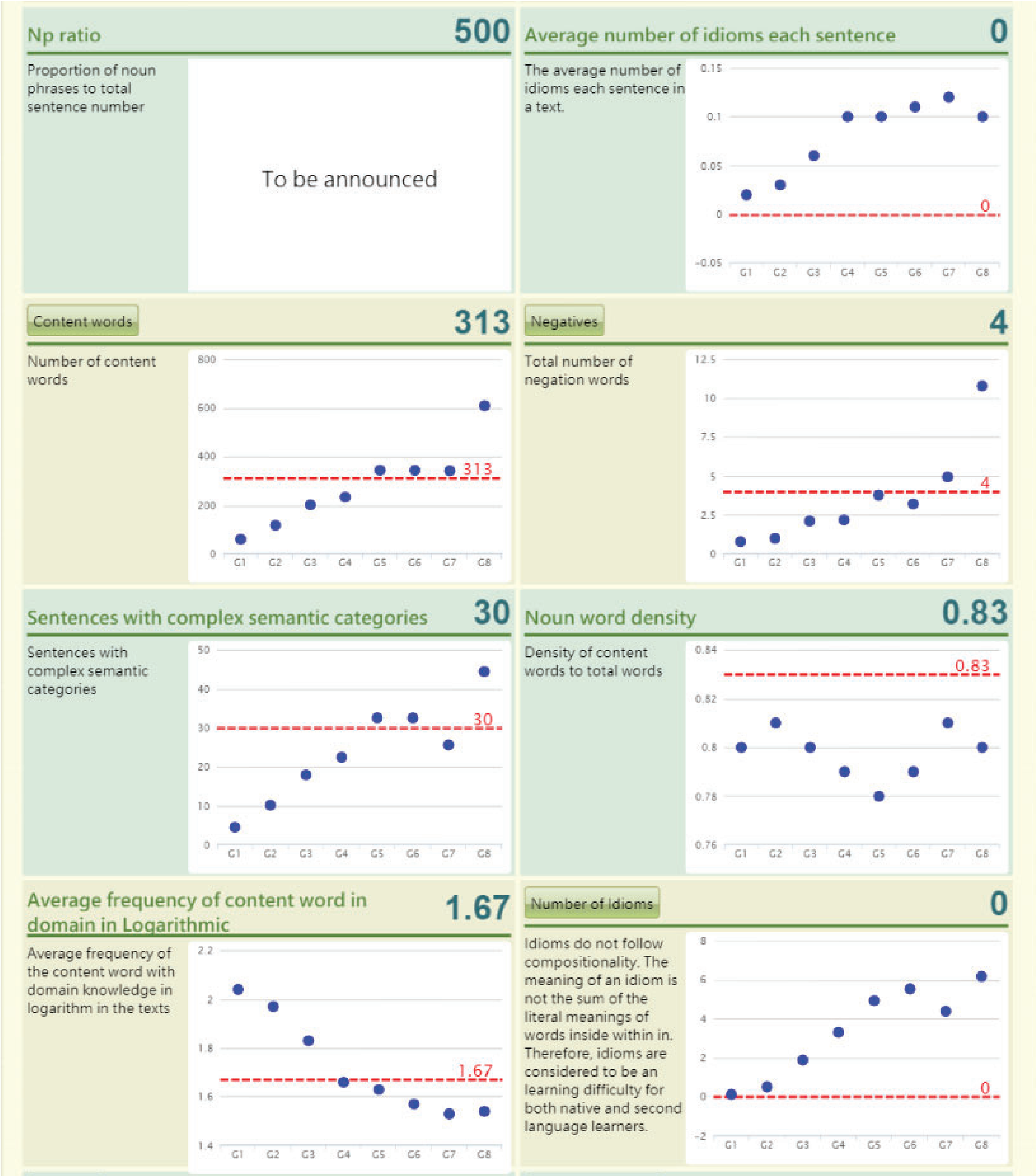
Video 4 text analysis



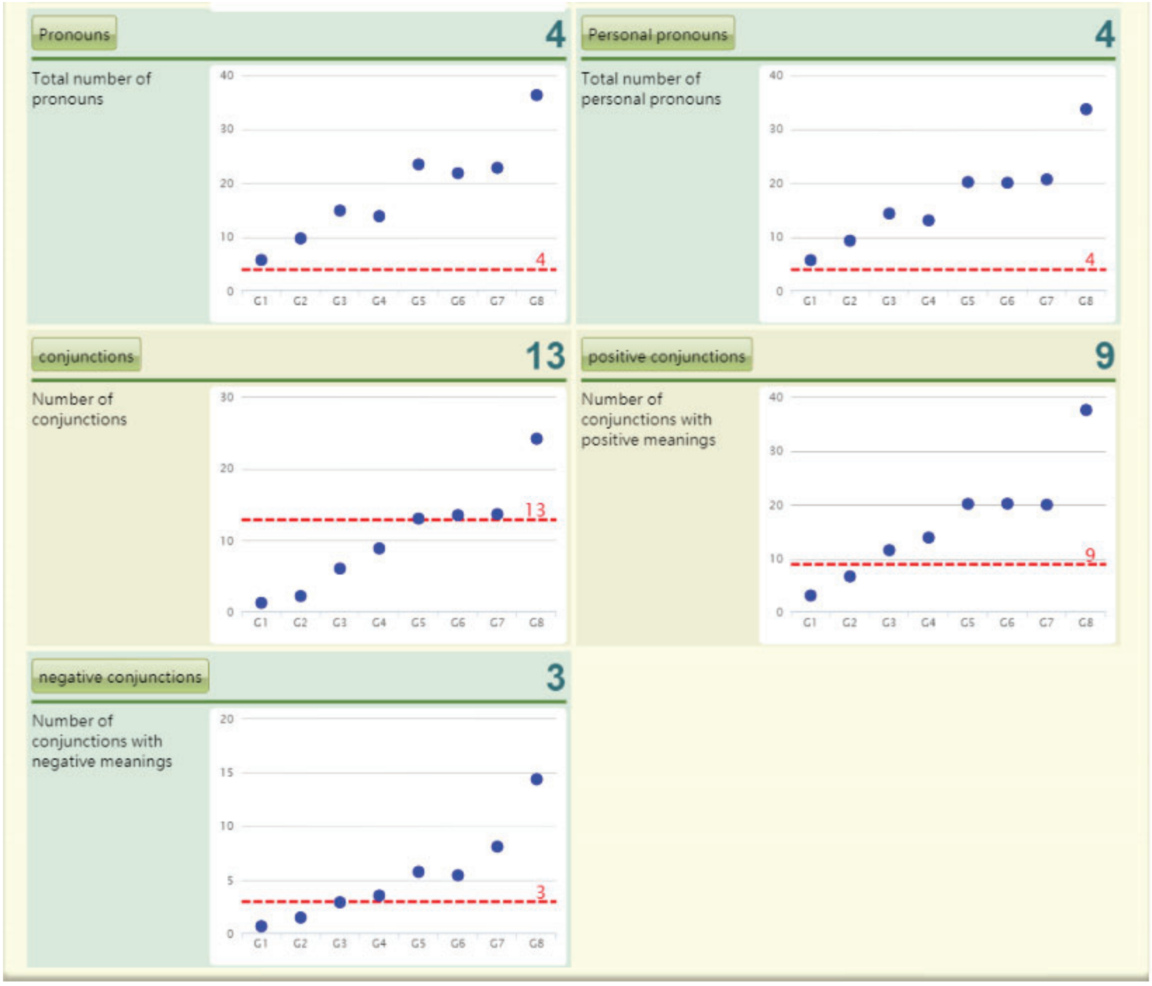
Video 4 text analysis v4.1



Video 4 text analysis v4.2

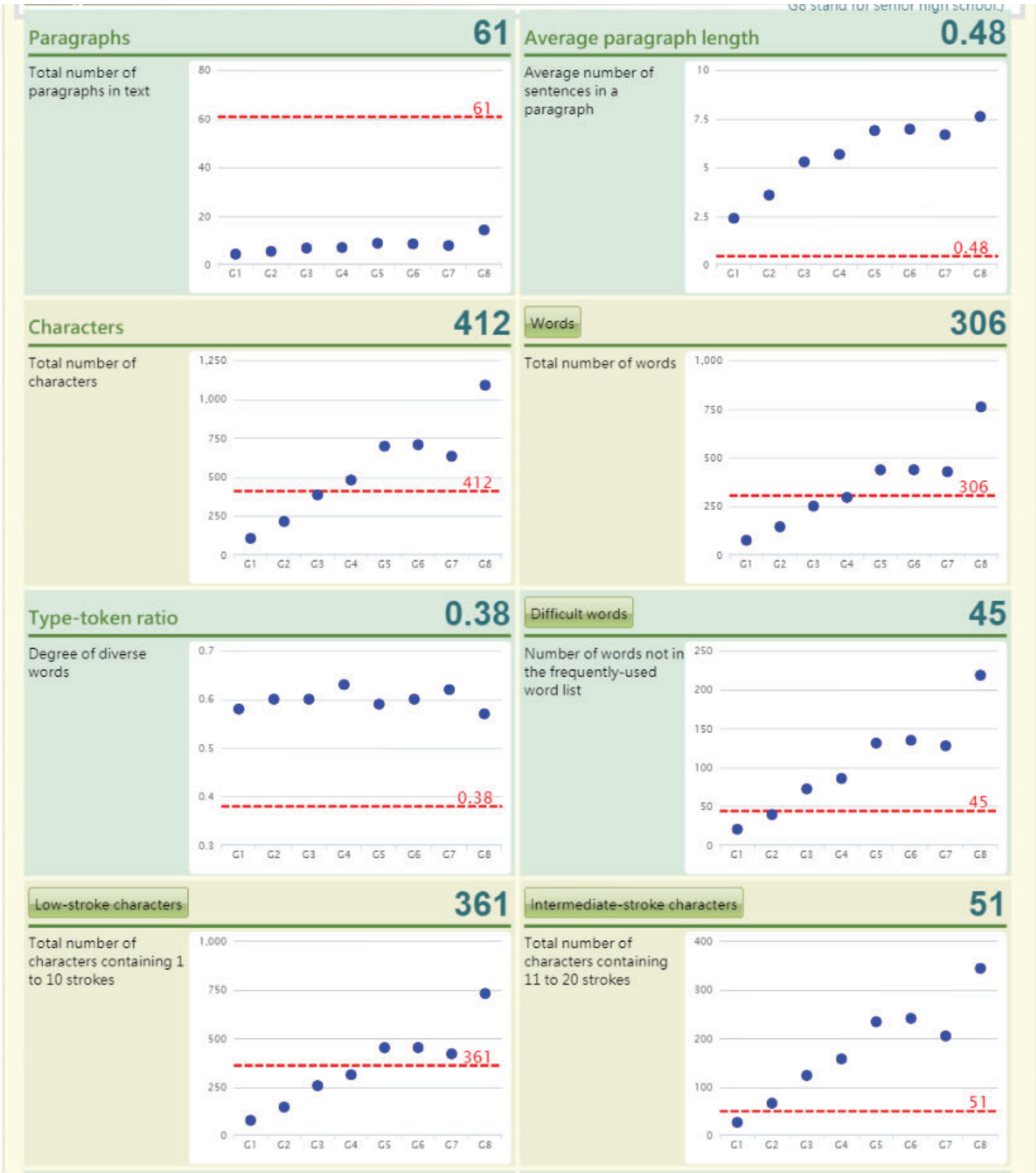


Video 4 text analysis v4.3

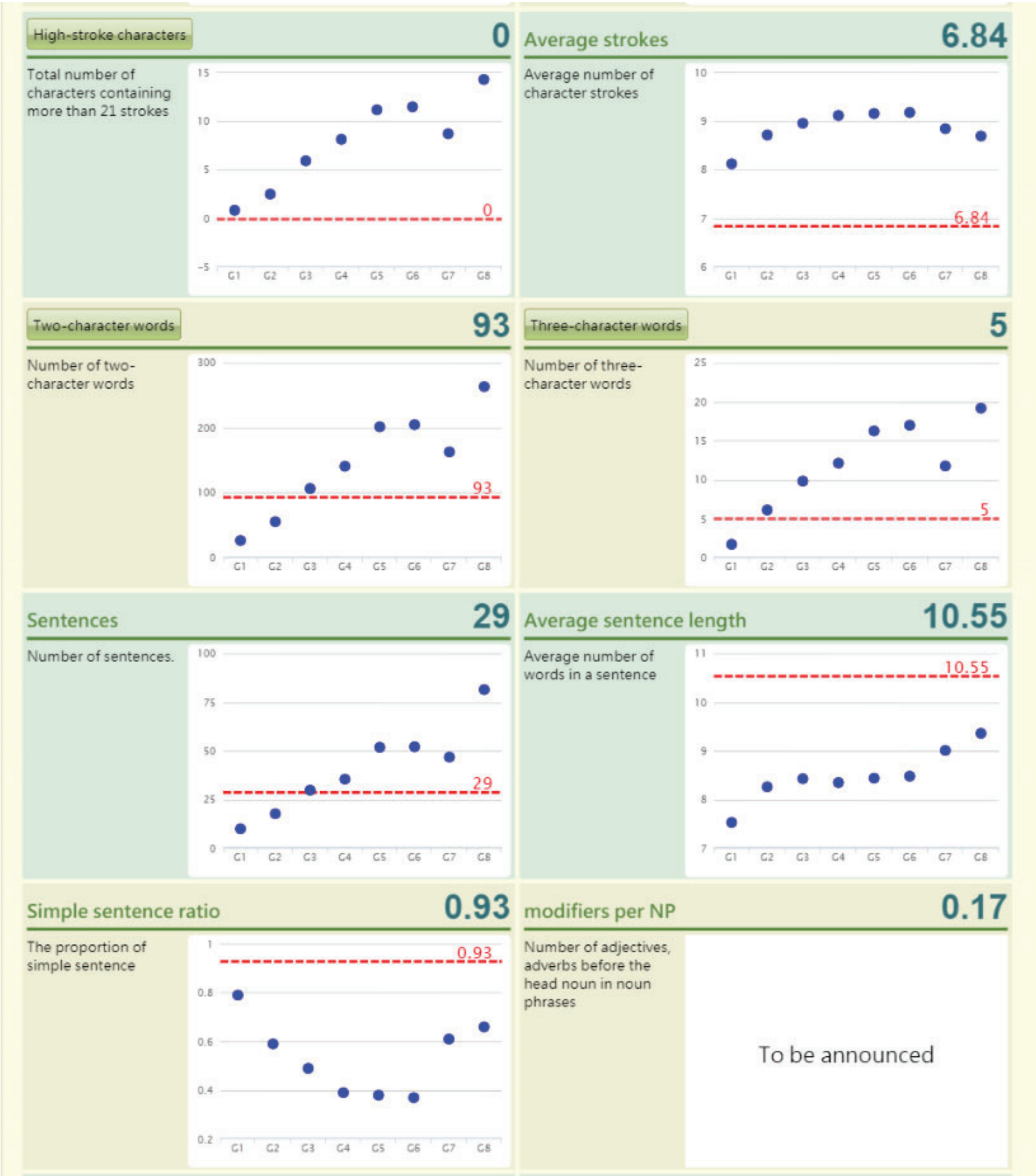


Video 4 text analysis v4.4

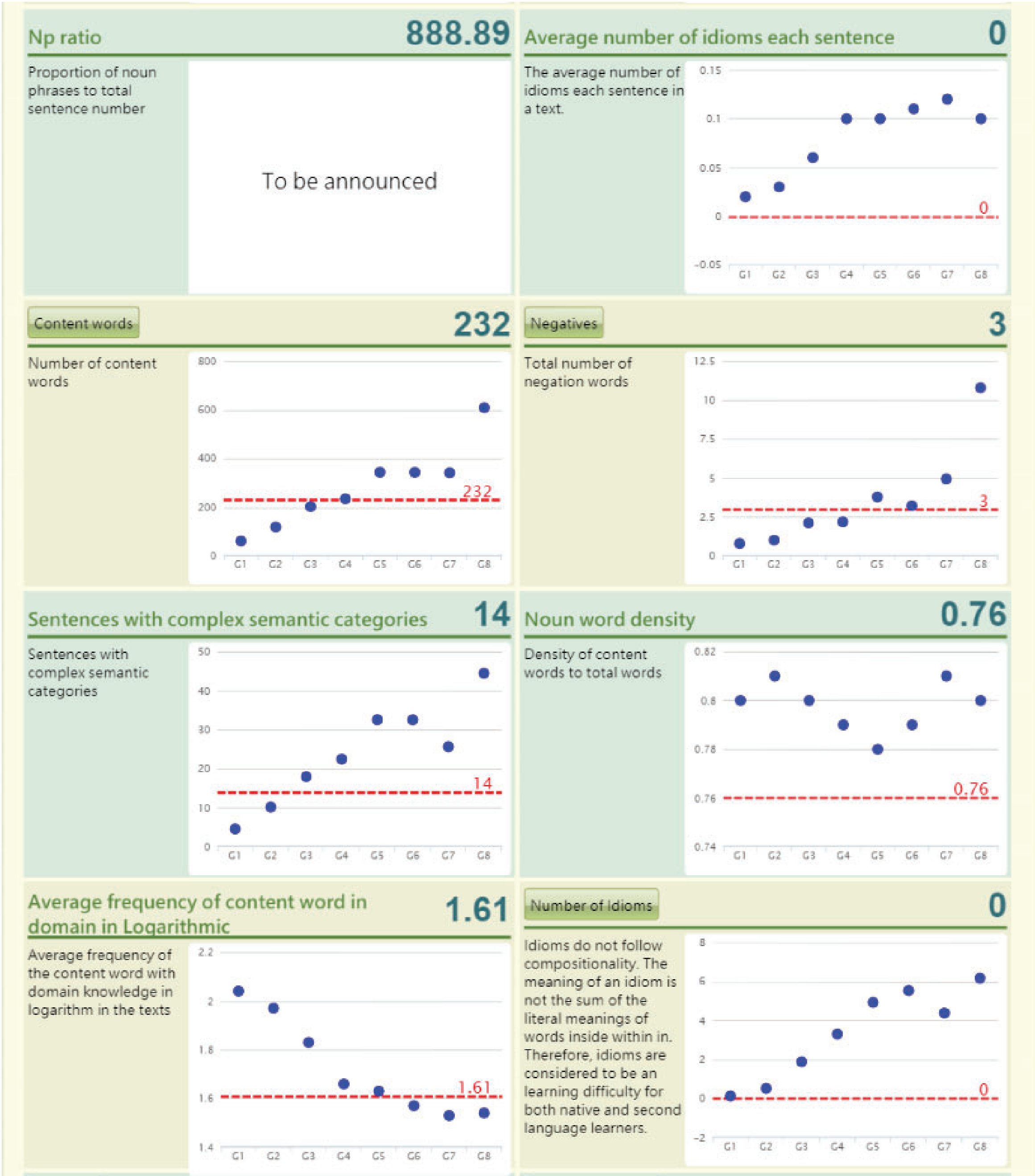
Video 5 text analysis



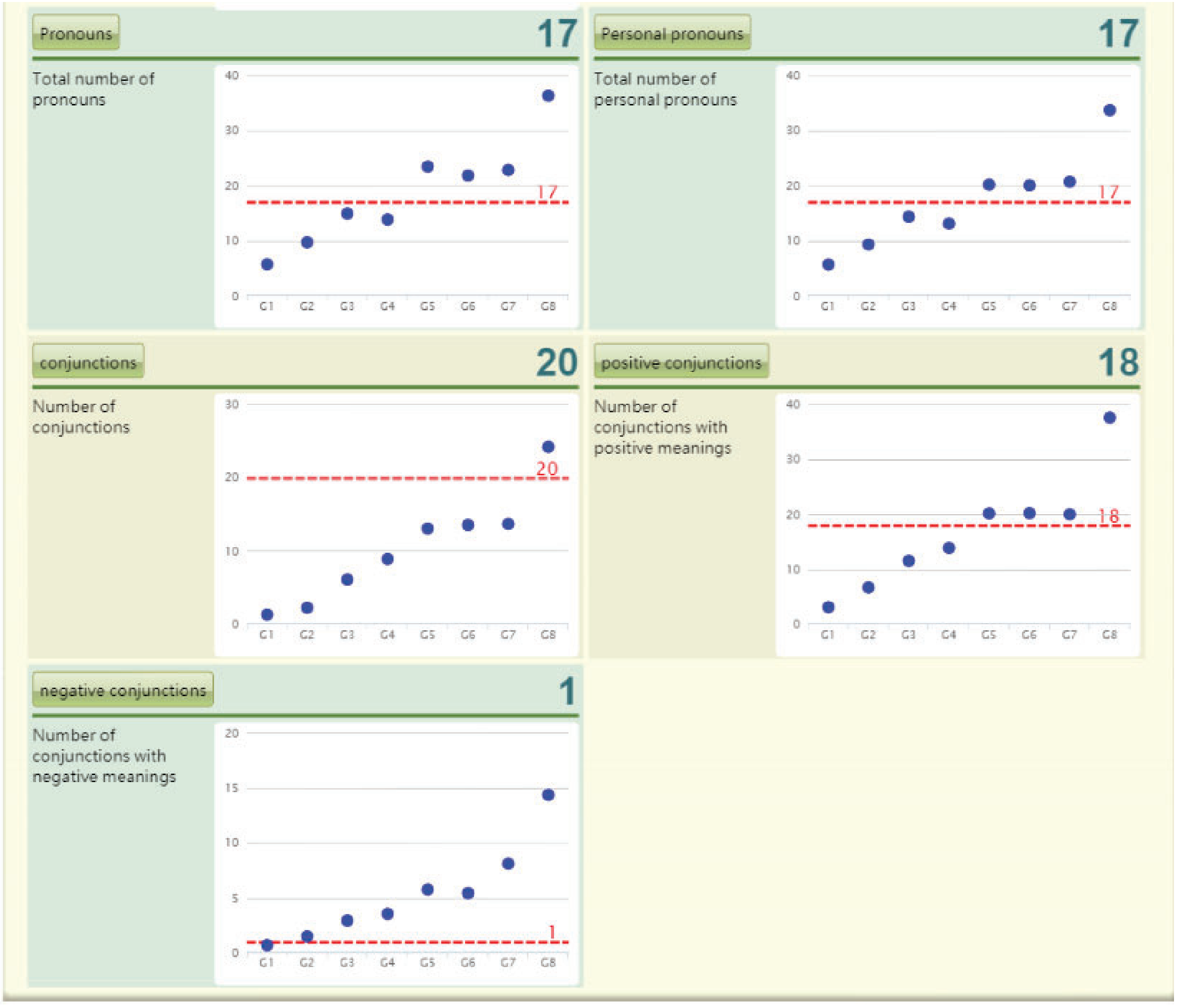
Video 5 text analysis v5.1



Video 5 text analysis v5.2



Video 5 text analysis v5.3



Video 5 text analysis v5.4

Appendix F

Raw comprehension scores for CS in Experiment 2

Participants	Video 1	Video 2	Video 3	Video 4	Video 5	Average
P01	3.00	5.00	2.00	4.00	3.00	3.40
P03	4.00	4.00	3.00	4.00	4.00	3.80
P05	2.00	4.00	2.00	4.00	3.00	3.00
P09	4.00	5.00	3.00	2.00	3.00	3.40
P10	3.00	3.00	3.00	2.00	4.00	3.00
P11	5.00	5.00	4.00	4.00	4.00	4.40
P12	4.00	4.00	5.00	4.00	4.00	4.20
P15	3.00	5.00	2.00	5.00	3.00	3.60
P17	3.00	4.00	2.00	4.00	4.00	3.40
P19	3.00	5.00	3.00	5.00	4.00	4.00
P21	4.00	4.00	4.00	3.00	2.00	3.40
P23	3.00	5.00	5.00	4.00	1.00	3.60
P25	5.00	4.00	3.00	2.00	3.00	3.40
P27	4.00	3.00	2.00	3.00	2.00	2.80
P29	4.00	5.00	3.00	4.00	4.00	4.00
P31	4.00	5.00	2.00	5.00	4.00	4.00
P33	3.00	4.00	3.00	3.00	4.00	3.40
P35	4.00	5.00	3.00	4.00	4.00	4.00
P37	3.00	5.00	2.00	4.00	4.00	3.60
P39	4.00	5.00	1.00	4.00	3.00	3.40
P62	5.00	4.00	3.00	4.00	2.00	3.60
P63	4.00	4.00	3.00	3.00	2.00	3.20
P67	3.00	4.00	2.00	4.00	2.00	3.00
P68	3.00	5.00	2.00	4.00	1.00	3.00
Average	**3.63**	**4.42**	**2.79**	**3.71**	**3.08**	**3.53**

Raw comprehension scores for ES in Experiment 2

Participants	Video 1	Video 2	Video 3	Video 4	Video 5	Average
P02	4.00	5.00	3.00	4.00	4.00	4.00
P04	3.00	4.00	2.00	4.00	3.00	3.20
P06	3.00	4.00	2.00	4.00	4.00	3.40
P07	5.00	5.00	2.00	4.00	4.00	4.00
P08	3.00	3.00	3.00	4.00	3.00	3.20
P13	2.00	4.00	4.00	2.00	3.00	3.00
P14	5.00	2.00	4.00	5.00	3.00	3.80
P16	4.00	4.00	3.00	3.00	2.00	3.20
P18	3.00	3.00	3.00	2.00	3.00	2.80
P20	5.00	5.00	4.00	5.00	3.00	4.40
P22	4.00	4.00	3.00	4.00	4.00	3.80
P24	3.00	2.00	0.00	2.00	2.00	1.80
P26	1.00	3.00	3.00	5.00	4.00	3.20
P28	0.00	3.00	2.00	4.00	1.00	2.00
P30	3.00	4.00	3.00	4.00	4.00	3.60
P32	1.00	3.00	3.00	5.00	4.00	3.20
P34	3.00	3.00	3.00	4.00	4.00	3.40
P36	3.00	3.00	1.00	3.00	3.00	2.60
P38	5.00	3.00	3.00	3.00	4.00	3.60
P40	3.00	4.00	2.00	4.00	4.00	3.40
P56	3.00	4.00	3.00	3.00	3.00	3.20
P57	4.00	5.00	2.00	5.00	4.00	4.00
P61	4.00	3.00	2.00	4.00	3.00	3.20
P64	2.00	3.00	1.00	3.00	1.00	2.00
P65	3.00	5.00	2.00	4.00	3.00	3.40
P66	4.00	4.00	3.00	4.00	2.00	3.40
P69	5.00	1.00	1.00	3.00	3.00	2.60
P70	3.00	3.00	2.00	0.00	2.00	2.00
Average	**3.25**	**3.54**	**2.46**	**3.61**	**3.11**	**3.19**

Raw comprehension scores for NS in Experiment 2

Participants	Video 1	Video 2	Video 3	Video 4	Video 5	Average
P41	2.00	4.00	2.00	4.00	3.00	3.00
P42	4.00	4.00	3.00	3.00	4.00	3.60
P43	5.00	4.00	2.00	4.00	5.00	4.00
P44	1.00	4.00	4.00	3.00	4.00	3.20
P45	2.00	3.00	2.00	1.00	2.00	2.00
P46	2.00	4.00	1.00	4.00	0.00	2.20
P47	2.00	2.00	2.00	4.00	2.00	2.40
P48	0.00	2.00	1.00	2.00	1.00	1.20
P49	4.00	5.00	4.00	2.00	4.00	3.80
P50	5.00	2.00	2.00	4.00	3.00	3.20
P51	5.00	4.00	2.00	3.00	4.00	3.60
P52	4.00	3.00	2.00	4.00	2.00	3.00
P53	3.00	4.00	3.00	4.00	4.00	3.60
P54	3.00	4.00	1.00	5.00	3.00	3.20
P55	4.00	3.00	3.00	3.00	3.00	3.20
P58	2.00	4.00	3.00	4.00	3.00	3.20
P59	5.00	5.00	5.00	4.00	4.00	4.60
P60	3.00	4.00	2.00	4.00	4.00	3.40
Average	**3.11**	**3.61**	**2.44**	**3.44**	**3.06**	**3.13**

Average comprehension scores for CS, ES and NS in Experiment 2

Group	Video 1	Video 2	Video 3	Video 4	Video 5	Average
CS	3.63	4.42	2.79	3.71	3.08	3.53
ES	3.25	3.54	2.46	3.61	3.11	3.19
NS	3.11	3.61	2.44	3.44	3.06	3.13

Appendix G (Ethics approval) of this thesis has been removed as it may contain sensitive/confidential content