Comparing the impact of subtitles on learning:

Automatically generated vs. corrected subtitles

Wing Shan Chan

BSc, University of New South Wales, 1996

GradDipMTh, University of Technology, Sydney, 1999

MT&I, Macquarie University, 2015

Faculty of Human Sciences

Department of Linguistics

Macquarie University

This thesis is presented as a partial fulfilment to the requirements

for the Master of Research

21st October, 2016

Table of contents

Chapter 3:

Abstract

Subtitles have been proven to benefit language acquisition, improve comprehension, listening skills and word recognition in an educational context. According to Cognitive Load Theory (CLT), information from different modalities that are presented simultaneously may cause cognitive overload due to the redundancy effect. However, research has shown an increase in learning when information is presented concurrently both visually and auditorily. The result can be explained by Dual Coding Theory: when redundant information is complimenting each other during processing, no cognitive overload results. Adding subtitles is essential for equal access as many online video lectures have become a major channel for learning in education. However, conventional subtitling is both expensive and time consuming, with the result that automated subtitling is a potentially powerful solution if the issue of accuracy and readability can be resolved. The focus of this study is to investigate the impact of automated and corrected subtitles on learning (tested in the context of micro-economic principles). A video was shown to participants, who were randomly assigned to 3 test groups: English lecturer with no subtitles, English lecturer with automated subtitles, and English lecturer with corrected subtitles. Participants were asked to complete a pre-test, an effort test and a post-test. Biographical information was collected for data analysis and interpretation. The objectives of this study are 1) to determine whether adding automated English subtitles and corrected English subtitles to English video improves learning; 2) to determine whether these subtitles impact differently; and 3) to compare the amount of cognitive load induced by automated and corrected subtitles on these groups.

Statement of Candidate

I certify that the work in this thesis entitled 'Comparing the impact of subtitles on learning: Automatically generated vs. corrected subtitles' 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.

Wing Shan Chan

Student ID: 43216625

21 October, 2016

List of figures

List of tables

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Associate Professor Jan-Louis Kruger, who has been a constant inspiration throughout my journey of thesis writing. Jan-Louis, your valuable advice, encouragement and guidance have made this research possible.

Thank you to all the lecturers who teach ECON111 in MUIC, especially Prashan, without your collaboration, I would not have had the data to work with in my project.

Thank you to Gordon for your help in data collection and professional input in data analysis, all the best with your PhD.

Thank you to Katherine: you have been a constant backup in the past four years. You have guided me through the darkest journey of my life. Your constant support and reassurance has helped me to identify who I really am and to keep my sanity intact to make this achievement possible; to Anna and staff from Brighter Future: your help to my family is so valuable, you have strengthened me in this process so I can achieve whatever I want to.

Last but not least, to Neil, your unconditional love and acceptance, nonstop support and encouragement are my main driving force. Thank you for all the intellectual exchange, delicious cooking, and your professional inputs in data coding. Most of all, thanks for believing in me, letting me fly and reach for my dream. Of course, to all my little angles and monsters: Mary, Katie, Emma and Sophie, you are the reason why I can do all these. You make me strong.

Chapter 1. Introduction

Subtitles have been proven to be beneficial in learning in a number of previous studies in different contexts. With the large numbers of educational institutions using online video lectures as a major part of their educational programs, the need for subtitled videos in education has also increased dramatically. This could potentially be extremely expensive, which means that automated subtitles are potentially powerful in fulfilling that need. However, this can only happen if current problems of accuracy, reading speed and text chunking can be resolved. Also, the impact of subtitle quality in an educational context has to be investigated more rigorously.

Various studies have proven subtitles to be beneficial in language acquisition, vocabulary recognition and comprehension in an educational context (cf. eg. Bird & Williams, 2002; Danan, 2004; Garza, 1991; Markham, 1999; Moreno & Mayer, 2002; Perego, Del Missier, Porta, & Mosconi, 2010; Vanderplank, 1988). Results of these studies also suggested the presence of a dual coding effect, namely that simultaneous presentation of visual and auditory information – subtitles and spoken dialogues – improves information processing, thus assisting language learning, recall and comprehension. Researchers are curious to find out the impact of subtitles on cognitive load and have to find answers to whether subtitle reading would cause the brain to overload cognitively.

Cognitive load theory (CLT) is a cognitive architecture based on the concept of the interaction between a limited working memory and a comparatively unlimited long term memory. According to CLT, information can be automated in cognitive schema and can then bypass working memory during mental processing. This can free up space in working memory (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). According to Paas and Van Merrienboer (1994), cognitive load can be defined as a multidimensional construct that represents the load that the learner needs in performing a certain task. CLT differentiates 3 types of cognitive load: intrinsic load, extraneous load, and germane load. Intrinsic load is an interaction between the

nature of the material that needs to be learned and the learner's level of expertise. Extraneous load results from badly designed instructions and poorly presented task. Germane load is related to the process that constructs and automates schemas (Paas et al., 2003). In order to know how subtitle reading impacts on cognitive load, researchers have employed several methodologies to evaluate cognitive load by measuring mental load, mental effort and performance. These methods include subjective rating scale techniques, physiological techniques, and task-performance-based techniques (Paas et al., 2003). Among these methodologies, physiological techniques, such as eye tracking and EEG, have the potential to provide online information and insight on cognitive processing of subtitle reading.

There is evidence (ref. Diao, Chandler, & Sweller, 2007; Mayer, Heiser, & Lonn, 2001), based on CLT, suggesting that subtitles might cause cognitive overload and thus decrease performance due to the redundancy effect. The redundancy effect occurs when learners have to coordinate mentally the same information presented concurrently in different channels (Diao et al., 2007). This action causes learners to split their attention between materials (Kalyuga, Chandler, & Sweller, 1999) which leads to cognitive overload and interferes with the learning of the intended information. In contrast to other studies, Kruger, Hefer, and Matthew (2013) found no redundancy effect in their study when information was presented in multiple channels at the same time, and they explained the result by dual coding effect. Their study showed that subtitles actually support learners in processing and understanding information. Other factors that could affect the impact of subtitles on a learner's outcomes include individual learning preferences, language proficiency, presentation of subtitles, instructional design, individual reading speed and literacy level.

With the proven benefits of subtitles and the increased use of online video lectures by large numbers of educational institutions, it is logical to seek for a solution to produce large quantities of subtitles economically within a shorter time frame. Automatically generated subtitles could be instrumental in meeting the need to create subtitles fast and affordably.

Automated subtitles are created by means of automatic speech recognition (ASR) software to provide a verbatim displayed transcript from spoken dialogues (Wald & Bain, 2008). The text transcription is then synchronised with the video by means of the timing information so that subtitles can be generated automatically (Wald & Bain, 2008). However automated subtitles can only be used reliably if their problems with accuracy, reading speed and text chunking can be solved. There are assessment models that are used to assess the quality of live captioning (Romero-Fresco & Pérez, 2015). The NER model is one that was introduced by Romero-Fresco in measuring the quality of live captioning, and the study was conducted in UK. According to Romero-Fresco and Pérez (2015), the accuracy rate of any live captions has to reach at least 98% accuracy in order to be considered acceptable in the NER model.

Previously, most studies were conducted to investigate the impact of subtitles on learning, usually comparing between subtitles and no subtitles condition. There are hardly any empirical studies that compare the type or the quality of subtitle conditions. With the knowledge that there are problems with automated subtitles, it is still uncertain whether automated subtitles have an impact on learning in comparison to video without subtitles. It seems logical that these automatically generated subtitles have to be corrected, but in the interest of efficiency, it is important to determine to what extent the correction of mistakes and grammar impacts on learning and cognitive load. It is the goal of this study to fill the gap of knowledge in this field of study, to find out if there is any difference in impact between unsubtitled video, video with automatically generated subtitles, and video with corrected automated subtitles on learning.

Based on the review of literature and the current standard quality of automatically generated subtitles, it is the interest of this study to find answers to the following research questions: 1) Does adding automatically generated or corrected subtitles to an educational video improve learning? 2) Do these subtitles impact differently on learning? 3) Is there a difference in impact of automatically generated and corrected subtitles on cognitive load? The objectives of

this study are therefore 1) to determine whether adding automatically generated English subtitles and corrected subtitles to English video improves learning; 2) to determine whether these subtitles impact differently; and 3) to compare the amount of cognitive load induced by automatically generated and corrected subtitles on these groups in comparison to unsubtitled video.

Chapter 2. Literature Review

2.1 Introduction

Subtitles are used extensively in movies and television programs. They were initially added to television programs targeted for the hearing impaired population, but it was observed later that the hearing population are also benefiting in various ways such as language learning, both in mother tongue and second language (Garza, 1991). Previous research has shown a difference in vocabulary acquisition, comprehension and task performance between viewers who saw video with subtitles and viewers who saw the same video without subtitles, in favour of the group that saw subtitled video. However, there is still not enough research to answer the question of how subtitles impact learning and under what conditions it has the optimal influence. As many educational entities use videos in their online learning programs, a large amount of subtitling is required to ensure equality of access. But creating proper subtitles is not only expensive, it is also very time consuming. Automated subtitles could be a potential solution to this. If problems with their accuracy, chunking of text and reading speed can be solved while creating automated subtitles, it could be a very powerful tool in improving learning in education. Automated subtitles are the text generated automatically with the help of speech recognition tools which is then synchronized with the video, and corrected subtitles are the same transcripts with grammar, spelling and other error corrected (see Figures 1.1 and 1.2).

Figure 1.1 (a, b) Automatically generated subtitles

Figure 1.1 (c) Automatically generated subtitles

Figure 1.2 (a) Corrected subtitles

Figure 1.2 (b, c) Corrected subtitles

The study will investigate the impact of both types of subtitling on performance and cognitive load. Although subtitles and subtitling are often used for translation subtitles when the subtitles are in a different language than the dialogue, and captions and captioning are sometimes used for same language subtitles for the Deaf and hard of hearing, the terms will be used interchangeably throughout this paper.

2.2 Background

Subtitles are widely known for entertainment in the context of foreign movies and television programs. It was actually introduced initially for the deaf and hard of hearing, and in the United States, for example, the National Captioning Institute (NCI) started the first television transmission of closed captioned programming in service to the hearing impaired population in 1980 (Garza, 1991).

Subtitling is a translation practice that translates the spoken form of language into written text (Garza, 1991), which usually appears on the lower part of the screen, and that presents the original dialogue of the speakers, as well as other elements that show in the image such as letters, inserts, graffiti, inscriptions and placards, and the information that is contained in the soundtrack such as songs and voice off . Same language subtitles are also known as intralingual subtitling, subtitling for the Deaf and hard of hearing or SDH, or captioning in certain contexts. It has also been labeled bimodal and unilingual in scholarly literature (Danan, 2004, p. 68). Same language subtitles are subtitles in which the language of the spoken dialogue is the same as that of the subtitles including sound effects and song lyrics that can be viewed by the audience.

As subtitles are increasingly used by a variety of users other than the hearing impaired population, researchers are asking questions about how subtitles can be used as a pedagogical aid in language learning, and also as a positive application in supporting students studying in foreign languages.

2.3 Subtitling as language support

According to the Australian Government Department of Education and Training (2016) , the number of international students enrolling in higher education in Australia has increased by 9.2% from 2014 to 2015. Students from China contributed the largest share of enrollments at 26.4% of all foreign students, and there was a 12.8% increase in Chinese student enrollment

between 2014 and 2015. Some other top nationalities that contributed to the total international students included India, Vietnam, Thailand, and South Korea. Even though international students are required to attain a certain level of language proficiency in IELTS (International English Language Testing System) or similar for any academic applications, students who have gained admission still experience language as their biggest hurdle that can potentially hinder their academic performance (Zhang & Mi, 2010). According to Zhang and Mi (2010), among these language problems, speaking and listening skills were reported to be particularly affected in the early years of their overseas study. Therefore, if subtitling could offer a tool for addressing some of these language issues, it could be a huge benefit to international students.

2.4 Impact of video in education

Compared to conventional text-books, information in audio-visual presentations improves comprehension and recall. Research has shown that the use of video in educational context has a positive impact on learning (Armstrong, Idriss, & Kim, 2011; Merkt, Weigand, Heier, & Schwan, 2011; Wilson, Park, Curtis et al., 2010), especially in patient education in a medical context, in which video was proven to be a more effective educational tool than printed materials on information that is procedural in nature (Armstrong et al., 2011; Wilson et al., 2010). Merkt et al. (2011) suggested that interactive videos, the ones that allow viewers to have control over the way they view the information such as pause, stop, rewind, skip, and fast forward, were comparable or even better than the conventional text-book. Their study indicated that interactive videos impact significantly more on learning than videos that are strictly medium controlled because of the self-regulated feature of information processing.

Video use in modern education both online and in class to establish learning engagment is crucial for positive learning. Subtitles has the effect of engaging students, and thus assist students to engage in the learning process because students look at the text on video naturally as an automatic human behaviour (d'Ydewalle & De Bruycker, 2007; d'Ydewalle, Praet, Verfaillie, & Rensbergen, 1991). As a large number of educational entities, such as Coursera,

Khan Academy, Academic Earth, and other open courses organized by Universities such as Stanford, MIT, Yale and Harvard, are using online video lectures as part of their educational programs, adding subtitles are essential to ensure people with different needs, including those with hearing loss and those with learning needs, can gain access to the content of the video lectures.

2.5 The benefits of subtitles

The benefits of subtitles in education and language learning have been proven in various studies (Bird & Williams, 2002; Danan, 2004; Garza, 1991; Markham, 1999; Moreno & Mayer, 2002; Perego et al., 2010; Vanderplank, 1988). In Garza's study (1991), he found that the use of subtitles could bridge the gap between reading and listening comprehension, which facilitated language use in proper context. The findings by Garza provide evidence to implementing subtitles in learning materials for L2 students during their early years of overseas study. As Zhang and Mi (2010) found that speaking and listening skills are particularly problematic in L2 students, subtitles could be beneficial in addressing some of these language issues. Likewise, Danan (2004) also indicated that subtitles enhanced listening comprehension of non-native language learners, and that they facilitate language learning through deepening cognitive processing.

Bird and Williams (2002) showed that exposure to subtitles increased word learning and word recognition which resulting in better comprehension. Vanderplank (1988) stated that the use of subtitles increased comprehension. Markham (1999) demonstrated that the availability of subtitles improved word recognition in University-level ESL students. He also noted that students with an advanced level of second language reading ability can use subtitles to develop their listening skill. In Moreno and Mayer's study (2002), results also supported that the presence of subtitles increased comprehension in a multimedia learning context. But they also mentioned that the benefits of subtitle exposure are only limited to situations where learners received no other visual presentation concurrently. In the early study on subtitles in

language learning, Vanderplank (1988) investigated the potential benefits of subtitles on 15 European ESL students from watching 9 hour-long sessions of television programs. The students watched the subtitled programs in the succession of 9 weeks with one hour program watching each week. Activities and discussion were carried out after the 9 hour-long program watching. Students reported that subtitles have made the English programs much easier to understand, and they were conscious of language learning during the process. There was a high level of language retention and recall from those used in the programs, and no cognitive overload was reported. Vanderplank's study (1988) has shown that when text was presented simultaneously with the audio presentation, it encouraged language learning with no evidence of cognitive loading.

Similarly, Perego et al. (2010) conducted a study in which a subtitled film excerpt was used for cognitive processing analysis. Eye movement data, word recognition and scene recognition were analysed and tested. The results have proven the hypothesis that presence of subtitles does not impede the cognitive processing of film or subtitles, viewers achieved a good level of word and scene recognition with no cognitive overload being detected. Subtitles have a positive effect on performance and understanding in the context of film appreciation, and were proven cognitively effective in the study of Perego et al. (2010). Moreno and Mayer (2002) later on conducted a more detailed study on the impact of subtitles in a multimedia context. Their research consisted of three experiments investigating the impact of adding subtitles to auditory explanation of the process of lightning formation. In experiment 1, students were presented with auditory only or auditory-visual modality explanation on lightning formation process. In experiment 2, simultaneous or a preceding animation were presented to auditory only or auditory-visual modality conditions, and results were compared. In experiment 3, environmental sounds were presented to auditory only or auditory-visual modality conditions. Results showed an improvement of comprehension when presenting words visually and by auditory means concurrently in explaining the lightning formation

process rather than just presenting auditory input. They noted that exposure to subtitles is only beneficial in situations where no extra visual presentation was presented simultaneously.

The results of these studies can be explained by Dual Coding Theory which suggests that combining visual and verbal information improves information processing (Paivio, 1991). These studies have shown that presenting verbal and visual information simultaneously assist language learning as well as recall and comprehension without causing cognitive overload. In other words, subtitles have proven to have a positive effect on learning and comprehension.

2.6 Subtitles and cognitive load

The impact of subtitles on cognitive load has been investigated in a few experiments. These studies intended to find answers to whether reading subtitles leads to an increase in cognitive load, and which methods can be used to reliably measure cognitive load. Cognitive load theory (CLT) is a cognitive architecture based on the concept of a limited working memory, with processing units for visual and auditory information that interact with a comparatively unlimited long term memory (Paas et al., 2003). According to CLT, information can be placed into a single unit in cognitive schemas that can be automated. Information then can bypass working memory during mental processing to free up space in working memory, and are thus not confined by the limitation of working memory (Paas et al., 2003).

Paas et al. (2003) stated that CLT identifies the concept of cognitive load as crucial in learning any complex cognitive task. They also indicated that the essence of the theory is in fact the instructional control of cognitive load to achieve transfer performance. It is known that the amount of working memory capacity allocated to a particular task largely impact the amount of learning and the complexity of what can be learned. If task demands exceed accessible cognitive capacity, and/or if cognitive resources are allocated inadequately, learning and complex task performance would be affected negatively.

Cognitive load can be defined as a multidimensional construct that represents the load imposed on the learner's cognitive system while performing a particular task (Paas & Van Merrienboer, 1994). According to Paas and Van Merrienboer (1994), the construct consists of two dimensions, a causal dimension and an assessment dimension, one that affects cognitive load while the other is affected by cognitive load. The causal dimension reflects the characteristics of task, or the characteristics of learner performing the task, or the relationship between both. The assessment dimension can be conceptualised as mental load, mental effort and performance which can all be measured. Regarding the causal dimension, task characteristics include format of task, complexity of task, pressure of time, and pacing of instruction; and learner characteristics consist of expertise level, age and spatial ability (Paas et al., 2003). Regarding the assessment dimension, mental load is an aspect of cognitive load that begins from the task and learner interaction. Mental effort is also an aspect of cognitive load referring to the cognitive capacity which reflects the actual cognitive load imposed on the learner by the task.

Performance is another aspect of cognitive load that can be defined by the learner's achievements (Paas et al., 2003). According to Paas et al. (2003), CLT differentiates 3 types of cognitive load: intrinsic load, extraneous load, and germane load. Intrinsic load is an interaction between the nature of the material that needs to be learned and the learner's level of expertise. It cannot necessary be influenced instructional designers. Extraneous load results from badly designed instructions. Germane load is related to the process that constructs and automates schemas.

It is clear in the model presented by Paas and Van Merrienboer (1994) that cognitive load can be evaluated by measuring mental load, mental effort and performance. A few methods have been employed regularly by researchers in previous studies including subjective rating scale techniques, physiological techniques, and task-performance-based techniques (Paas et al., 2003). There are certain assumptions in each of the techniques that measure cognitive load.

According to Paas et al. (2003), 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 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. In psycholinguistic research, electroencephalography (EEG) is used to study event-related potentials (ERPs), functional magnetic resonance imaging (fMRI), and specific measurements obtained by eye tracking, specifically pupillometrics, provide ways to give objective measurements of language cognitive processing directly, with the extra advantage of delivering real-time measurements (Kruger, 2016) According to Kruger (2016), eye tracking is the only method in physiological techniques that has been used largely in researching subtitling. Kruger (2016) indicates that some of the main foci in the studies of measuring eye movement include the way one and two-line subtitles are read differently, the distribution of attention between subtitles and the rest of the screen, and the effort required in subtitle reading. Other studies attempted to measure reading behaviour by analysing fixation duration through investigating the processing of native language (NL) and foreign language (FL) subtitles (Bisson, van Heuven, Conklin, & Tunney, 2014), and the impact of text chunking on subtitles processing and subtitles reading (Perego et al., 2010; Rajendran, Duchowski, Orero, Martínez, & Romero-Fresco, 2013).

Based on cognitive load theory (CLT), there is evidence (Diao et al., 2007; Mayer et al., 2001) suggesting that subtitles might cause cognitive overload and thus decrease performance. Mayer et al. (2001) proved the redundancy effect in the context of native English speakers. They found that students performed worse in terms of retention and transfer when subtitles were added, which caused students splitting their attention between different visual sources. Diao et al. (2007) also found redundancy effect in their study in the context of English as foreign language, when auditory materials were presented simultaneously with text presentation, even though comprehension improved immediately after viewing but poorer performance were observed in the succeeding auditory passage. They ascribe the decrease in performance to the impact of redundancy effects since the participants have to process information from different sources which interfered with their learning.

2.7 Verbal redundancy

According to Diao et al. (2007, p. 239), the redundancy effect is one of the instructional procedures generated by CLT, and it occurs when learners have to coordinate mentally the same information presented simultaneously in different channels. This action of having to process the same information in multiple channels causes learners to divide their attention between materials (Kalyuga et al., 1999) which leads to cognitive overload, and interferes with the learning of the intended information being presented. Kalyuga, Chandler, and Sweller (1998, p. 2) consider the redundancy effect to be less severe when redundant material is eliminated than included in a multimedia context. In other words, redundant information interferes with learning rather than facilitates it (Kalyuga & Sweller, 2014). Kalyuga and Sweller (2014) described that, within CLT, the redundancy effect resulted when redundant information from different sources are processed simultaneously, causing extraneous cognitive load to increase. The limited working memory is allocated to coordinate unnecessary information, thus decreasing the cognitive capacity for learning. They suggested

that redundant information that is not needed for learning should be omitted to avoid negative learning outcomes as a result of the redundancy effect.

In contrast to other studies that evidenced the redundancy effect, Kruger et al. (2013) found no significant impact on cognitive overload with or without subtitle exposure. In their study, students' performance was not affected with the double exposure of auditory and written form of content, and redundancy effect did not occur. Kruger (2013) explained that there is evidence of dual coding effect that the simultaneous presentation of verbal content in both auditory and text form may be free from the redundancy effect under certain conditions (Kruger, 2013, p. 36). He furthered argue that subtitles have positive impact on language performance as an overall outcome if no overly complicated multimedia materials are presented at the same time (Kruger, 2013, p. 35; 2016). Other variables such as the proficiency level of learners in the foreign language and the amount of cognitive load in different modalities also contribute to the extent of the impact of subtitles (Markham, 1999).

However, the findings of Kruger et al. (2013) showed a significant difference in cognitive load (CL) between groups with no subtitles producing higher CL and higher frustration levels. They concluded that their results suggested ESL students learning through the medium of English using same-language subtitles could experience lower CL if they were to use subtitled video rather than unsubtitled video. In other words, their study showed that the presence of same-language subtitles provided students with support that assist their processing and understanding of the learning content. The study by Kruger et al. (2013) was the first empirical research to measure CL directly in the presence of subtitles using EEG measurement and PCPD (percentage change in pupil diameter) in addition to comprehension test and self-report mental effort test. Their research findings, although limited in terms of the EEG results which relied on proprietary software rather than raw EEG data are important to the current study because both have a similar experimental design and procedure, and it evidenced the importance of further research on the impact of subtitles on cognitive load.

Homer, Plass, and Blake (2008) found that individual learning preferences – visual or verbal – together with other variables in the multimedia learning environment also have an impact on cognitive load. Participants with higher scores in visual preference have to apply greater cognitive effort to learn in the condition with slides and audio only narration (auditory modality) of the lecture, whereas participants with lower scores in visual preference have to apply greater cognitive effort to learn in the condition where a video of the lecture and slides (visual modality) were presented at the same time. Even though there is a significant difference in cognitive load associated with the video, there is no significant difference in the learning outcome in all the tested conditions where information from different modalities were presented simultaneously. Based on research outcome, Homer et al. (2008) suggested that for instructional techniques to be efficient, learners' individual differences are required to be taken into account.

There are also other factors that affect the impact of subtitles on the learner's learning outcomes. These factors include the way information is presented, such as whether it is presented at a readable speed or in clear visual form, impacts on how much information is processed. Also the instructional design can affect how people process information. A clear and simple instruction is definitely easier for learners to process than a complex and ambiguous instruction. Other factors such as individual reading speed and literacy level can affect the processing of subtitles as well.

2.8 The availability of subtitles

There are barriers that make subtitles difficult to be available in every single audiovisual program despite of its benefits. In order to understand this difficulty, it is essential to discuss the process of subtitling and automatic subtitling.

Conventional subtitles are created by subtitlers who transcribe the audio dialogues into written text and add the text onto the screen using subtitling software, often editing the text

down to comply with a reasonable presentations speed that can be read easily by viewers while also taking in the visuals.

The Code of Good Subtitling Practice compiled by Ivarsson and Carroll (Diaz-Cintas & Remael, 2007) is an example of the general guidelines to ensure high quality, accurate and readable subtitles. There are some basic principles that are recommended to follow strictly in creating quality, accurate and readable subtitles (Diaz-Cintas & Remael, 2007). Diaz-Cintas and Remael (2007) stated that there are time and space constraints during subtitling. In displaying each chunk of subtitles, a minimum display of 1 second with short subtitles and a maximum display of 6 seconds are recommended for full two-liners on screen. In terms of space constraint, there should be no more than two lines per screen and the maximum characters, including spaces, should be between 33-35 on each line, some may allow up to 39-41 characters depending on the guidelines and software used. They also recommended all subtitles should be positioned in the center of the screen. Diaz-Cintas & Remael further stated the subtitles should be synchronised in time with the utterances, visual image and auditory signal, carefully timed with the cuts of each shot to avoid confusion, and subtitles should allow sufficient time for the audience to read and be able to process other semiotic signals simultaneously. Subtitles should be integrated as part of the AV text naturally. Diaz-Cintas and Remael (2007) also indicated that text reduction is essential during subtitling. They advised that subtitles may need to reduce as much as 30% of the original transcript and this process involved reformulating, condensing and adapting from the original text. Diaz-Cintas and Remael (2007) noted further that simplification of syntax is also necessary due to the space constraint of subtitling, unnecessary repetition, false starts and unnecessary interpersonal markers that can be omitted.

The conventional way of subtitling is considered expensive and time consuming. Professionals are needed in the process of creating quality subtitles, which are not always practical and possible. Even though professionally created subtitles are more accurate with

better quality, it has also become a major barrier and consideration to add subtitles in every video. Automated subtitles could be the solution in this aspect and could be widely used to reduce both cost and effort.

Automated subtitles are created by means of automatic speech recognition (ASR) software to provide a verbatim displayed transcript from spoken dialogues (Wald & Bain, 2008). The text transcription synchronises with the video by the timing information so that online videos can be subtitled automatically (Wald & Bain, 2008). Automated subtitles are easy to access, and they can be added to any education videos by just following simple steps of instructions while expert skills are not necessary. In addressing the challenge of subtitling (captioning is the original term used on the Google site) for users who upload their videos online, Google combined its automatic speech recognition (ASR) technology with the YouTube caption system to offer automatic captions, or auto-caps for short (Harrenstien, 2009). According to Harrenstien (2009), auto-caps automatically generate captions for video by using the same voice recognition algorithms also used in Google Voice. Google claimed YouTube auto-caps has an 80% accuracy rate (Berdichevsky, 2014; Blake, 2015; Bond, 2014b; Lockrey, 2015).

Even though it is far less accurate compared to the acceptable 98% accuracy rate set by the NER model (Romero-Fresco & Pérez, 2015), auto-caps is still helpful in a way that viewers can understand what was presented in the text. Furthermore auto-caps will improve with time as technology advances. In addition to automatic captions, Harrenstien (2009) wrote that Google launched automatic caption timing, or auto-timing, to make creating manual captions significantly easier. With auto-timing, only a simple text file is needed to be created for the transcribed text, then the captions will be created automatically by using Google's ASR technology and no special skills are needed in the process. By creating captions automatically, the barriers for video owners to add captions will considerably decrease, as time and resources to create professional caption tracks will reduce significantly. Automatic subtitling can save

large amount of resources for many educational entities in producing large quantities of videos for online distant learning programs (University of Washington, 2016).

2.9 Problems with automated subtitles

There are 3 main problems with automated subtitles: accuracy, reading speed and chunking of text. Typically, ASR has an average accuracy rate of 60-70% (Bond, 2014a), but the accuracy rate can improve up to at least 98% with editing, pre-recorded transcript, and the training of the computer to recognize the speaker's voice (Wald & Bain, 2008). Automatically generated subtitles can only be created with high accuracy under certain criteria (Jurafsky & Martin, 2009). According to Jurafsky and Martin (2009), these criteria are as follows: slow, clear and consistent speech style where the speaker speaks in a standard dialect, no background noise or sound effects, and limited vocabulary recognition at a time. However, it is hardly possible to have these criteria achieved in real life settings.

There are models of assessment that can be used in assessing live captioning quality (Romero-Fresco & Pérez, 2015). The US National Institute of Standards and Technology uses the following formula to calculate word accuracy (Dumouchel, Boulianne, & Brousseau, 2011):

$$\text{Accuracy rate} = \frac{N - \text{Errors}\ (D + S + I)}{N} \times 100 = \%$$

According to Romero-Fresco and Pérez (2015), this model draws on the basic principles of WER (word error rate), which have traditionally been applied in analysing the accuracy of speech recognition. In this model, N is the total number of words spoken by the speaker. The errors include D for word delete, S for substituted or I for inserted incorrectly by the speech recognition software. Romero-Fresco and Pérez (2015) noted that this model of assessment is for the use of speech recognition, and is not adaptable to European countries in which

captions are mostly created by respeakers, thus the model does not account for editing errors. In 2011, Romero-Fresco introduced the NER model, which draws on the basic principles of WER and factors in different severity of errors, at the same time accounting for different degrees of reduction in the subtitles, and emphasising the need for human intervention in assessing the caption quality (Romero-Fresco & Pérez, 2015). The following formula is the NER model to access the accuracy of live captions (Romero-Fresco & Pérez, 2015):

$$Accuracy = \frac{N - E - R}{N} \times 100$$

CE (correct editions):
Assessment:

In this model, N is the number of words in the subtitles, E stands for edition errors by strategies applied by the subtitler, and R stands for recognition errors by mispronunciation or mishearing from speech recognition or stenography. According to Romero-Fresco and Pérez (2015), captions must reach a 98% accuracy rate to be considered acceptable with the NER model. However, it is not the mere accuracy rate that determines the captions quality in the NER model, but the assessment that includes issues on the speed and delay of the captions, how the captioner deals with the audio speech rate, the overall flow of the on screen captions, speaker identification, the audiovisual coherence between original video and captions, and the time lost in the corrections. In Romero-Fresco's 2 years study in measuring the quality of live captions on UK TV, the broadcasters and the external reviewers analysed a total of 35 hours of live TV material from 216 programs, including roughly 362,000 words and 55,000 subtitles. The results of his study showed an average of 98.33% accuracy rate in live captions across most genres of programs, except talk shows and entertainment programs where there is very high speaking rate by the speakers that cause the decrease in accuracy.

Reading speed is another problem that affects viewers in comprehending the video content. Automated subtitles are usually presented at a faster speed since they are fully verbatim by using ASR. This problem is mainly caused by the high speech rate of the speakers in video

(Romero-Fresco, Forthcoming). Romero-Fresco (Forthcoming) concluded in his study that fully verbatim captions are not desirable especially with the introduction of the hybrid mode, which combines pre-recorded transcripts and live captions, because it is far too fast for any viewers to read and comprehend despite its high accuracy rate. In this aspect, reduction of text is necessary to allow readable and comprehensible subtitles, and inevitably professionals have to be involved in this process.

Lastly, research shows that text chunking increases the effectiveness of subtitle reading by reducing the amount of time spent on subtitles (Rajendran et al., 2013). Rajendran et al. (2013) investigate subtitle segmentation in grouping a block of text into coherent segments, also called text chunking, using eye-movement metrics to measure fixation durations. Their eye-tracking data shows that chunking the text by phrase or by sentence reduces the amount of time viewers spent on reading subtitles, and the processing of subtitles is easier through this presentation of the text.

2.10 Current practices

There are a few ways in creating subtitles on YouTube currently. The first step is to create a transcript of the spoken dialogues of the video in plain text (.txt) file. Through video manager, upload the .txt file to YouTube, which then turns the transcript into subtitles by providing automatic timing. This is done by automatically matching the audio of the video with the text of the transcript using YouTube's voice recognition technology (Berdichevsky, 2014; YouTube, n.d.). The second way of adding subtitles on YouTube is to type the text directly into the transcript text box through YouTube video manager. The video will be paused automatically as the text is typed in the transcript box. There are other function buttons such as pause, play and rewind for more accurate transcription. YouTube allows the transcript to sync with the video by just clicking the Set Timing button. The final synced subtitles will be published to the video when the procedure is completed (Berdichevsky, 2014; YouTube, n.d.). Another way of creating subtitles on YouTube is to let YouTube create the subtitles for the

video automatically using its speech recognition technology. The automated subtitles will be published on the video automatically if they are available. The automated subtitles can also be reviewed, edited or unpublished anytime it is needed (YouTube, n.d.).

The online open courses from different education entities have different ways of providing subtitles, mostly through the community volunteering in creating subtitles. In 2014, Coursera announced that their courses will be subtitled and translated by the Global Translator Community (GTC) which is a community of volunteers and partner organizations that work together to make educational content accessible to learners across the world (Coursera, 2014, n.d.). MIT OpenCourseWare also depends on volunteers to create and translate subtitles (MIT OpenCourseWare, 2010), while Khan Academy creates their own English subtitles professionally (Khan Academy, n.d.). Other open courses by Universities have not stated clearly how they create their subtitles, but most of their online lectures can be viewed on YouTube with closed captions that are quite accurate, but mostly verbatim.

2.11 Goal of this study

Previous studies evidenced the benefit of subtitles in learning. However, large quantities of subtitles are necessary for educational institutes to increase the accessibility of their online education programs. These institutes are facing the issue of creating large quantities of professional subtitles which are both expensive and take a lot of time to produce. Automated subtitles are potentially a solution in reducing time and cost while providing subtitles to online video lectures. With the current technology, automatically generated subtitles can only be produced with high accuracy under certain criteria, such as slow, clear and consistent speech style and speaking in a standard dialect, no background noise or sound effects, and limited vocabulary recognition at a time. However, this is not the case in most of the recorded videos in real life. Most studies conducted in the past were to investigate the impact of subtitles on learning, usually comparing the impact between subtitles and no subtitles condition. There are hardly any empirical studies that compare the type or the quality of

subtitle conditions. Knowing that there are problems with automated subtitles especially on its accuracy, it is uncertain if automatically generated subtitles have an impact on learning when compared with unsubtitled video.

Chapter 3. Method

3.1 Introduction

In order to address the issue on the impact of automated and corrected subtitles on learning and cognitive load, it is essential to answer the following research questions: 1) Does adding automatically generated or corrected subtitles to an educational video improve learning? 2) Do these subtitles impact differently on learning? 3) Is there a difference in impact of automatically generated and corrected subtitles on cognitive load? To answer these questions, this study was designed to conduct an experiment in an academic context where students were exposed to three versions of the same online lecture: 1) video with no subtitles; 2) video with automated subtitles, and 3) video with corrected subtitles. A pre-test and a post-test were carried out to determine the impact of the automated subtitles, corrected subtitles and unsubtitled video on learning. A cognitive load test was conducted to determine how each type of subtitles impact on cognitive load.

Base on the literature, the hypotheses of the study are: 1) automatically generated and corrected subtitles would result in better performance when compared to unsubtitled video; 2) automatically generated and corrected subtitles would result in lower cognitive load when compared to unsubtitled video; and 3) the corrected subtitles would outperform the other 2 test conditions.

3.2 Population and sampling

An availability sample of first-year students from the Macquarie University International College (MUIC) was used. The MUIC diploma is a very structured intensive program with students completing two courses at a time in 6-week terms. Due to the delivery format and the intensity of the program, student engagement and participation is high. In addition, all the students in Business and Economics programs have to complete an introductory course on the Principles of Micro-Economics and a high percentage of the students are from a non-English

background. As such, the students in the MUIC program constitute a sample of the wider population of first-year students in Business and Economics programs at the university. Furthermore, due to the fact that this is such a structured program, all the students in the program could be used in the experiment, provided they gave their informed consent in compliance with the ethics approval for the study.

The sampling of this group, and the fact that the topic of the lecture used in the experiment was aligned with the curriculum in the course, contribute to the ecological validity of the study. However, the sample of participants in this study was not tested for its homogenous base on language proficiency due to time limitation, and this could be done in future study. There were 7 groups ranging from 13-26 students per each group. A total of 141 students participated in the study and each group was randomly assigned to one of the 3 video conditions – English video without subtitles (E), English video with English automatically generated subtitles (EA), and English video with English corrected subtitles (EC). One group had to be discarded because the lecturer did not comply with the request to allow the experiment to be conducted at the beginning of the session, and as a result delivered the lesson content on the same topic before the video was shown. This invalidated the data for this group who had an additional input on the topic which means that their performance in the post-test could not be compared to the groups that did not have the same input. Other participants were excluded due to incomplete responses (for example a failure to complete either the pre-test or the post-test). Only 92 sets of data could therefore be used in the study due to either incomplete data or invalid study procedure. The final division between the groups was 21 students in condition E, 34 students in condition EA, and 37 students in condition EC.

The students were mainly from China, India, Vietnam, South Korea, and Thailand, although there were also a number of local Australian students. A multiple regression analysis was performed with self-reported cognitive load and performance on pre-test, post-test, and full

post-test scores as dependent variables, and months students spent in an English-speaking country as predicting variable. The analysis revealed the length in an English-speaking country not to be a predictor of either cognitive load or performance, which meant that all 92 sets of data could be used.

3.3 Material

3.3.1 Video and subtitles

A video downloaded from MIT OpenCourseWare with the topic Elasticity of Supply and Demand from a course on micro-Economics was used (MIT OpenCourseWare, 2012). The topic covered in the lecture included a section on the shape of the supply and demand curve, a discussion on elasticity and inelasticity and how that shows on the supply and demand curve, the factors that affect elasticity of goods, how to clarify and distinguish between correlation and causation, and the shift in demand and how that affects price. The video was chosen because it was a recording of a real life lecture on Economics with the lecturer having a lesson and interacting with students. The video showed mainly just the lecturer on camera, he talked quite fast. Occasionally he wrote on the blackboard and drew diagrams. He also interacted with the students by asking them questions and they answered back. In the automated subtitles, where speech recognition software was used, the spoken dialogues between lecturer and students were not recognised due to the volume level and the microphone reception.

For the experiment, a clip of 25 minutes was selected with the topics above from a 50-minute video. The decision to use a shorter video was taken for practical reasons since the time allowed for taking down the data was very limited due to the structured curriculum. The video was then presented in three formats to three different groups. One video was unsubtitled, one with English subtitles that were automatically generated, and one with English subtitles that had been corrected.

3.3.1.1 Subtitle characteristics and quality

The video downloaded directly from MITOpencourseWare only provided the corrected version of the text transcript, which was uploaded and sync with the video through YouTube's automatic timing to create the subtitles. The transcript was possibly produced by volunteer because according to documentation, MIT only provided professional transcription after year 2012 (Khesin, 2012). The automated subtitle of the same video was available through YouTube. It was automatically created using YouTube's voice recognition technology that transcribed the spoken dialogues into text, then automatically matching the audio of the video using automatic timing to create the automated subtitles (Berdichevsky, 2014). The transcripts of both versions of the subtitles were manually inspected and analysed along with their presentation speed. The details of what has been found are shown in Table 3.1. It is evident from the analysis that the presentation speed of both versions is too high in many instances. The reason for the content of the corrected subtitles not being properly edited to reduce presentation speed may due to the fact that it was not created professionally at the time the video was produced.

According to the literature, an average reader can comfortably read text written on two full liner subtitles in six seconds, with a maximum of 37 characters in each line, which is a total of 74 characters (Diaz-Cintas & Remael, 2007). This is known as the six-second rule. According to the account given by Diaz-Cintas and Remael (2007) on the reasoning behind the six-second rule, since two frames enable one subtitle space, given that a projection of 24 frames for each second is needed for every film image, this gives subtitlers 12 subtitling spaces in each second. So in six seconds, a total of 72 subtitling spaces are available. The implication of the calculation is a low reading speed of 140-150 word per minute (wpm) or 11.67 CPS by conversion, and the new norm of reading speed increased to 180 wpm or 15 CPS for DVD (Diaz-Cintas & Remael, 2007). The average subtitling speed varies according to different guidelines. According to Romero-Fresco (2009), the official Ofcom's guidelines for UK

recommended a maximum subtitling speed of 160 wpm - 180 wpm, which converts to 13.33 CPS - 15 CPS; the French standard recommended subtitling speed of 120 - 150wpm, which converts to 10 - 12.5 CPS; and the Spanish standard of the recommended subtitling speed of 150 wpm, which converts to 12.5 CPS. Comparing the different guidelines of average subtitling speed standard, the subtitle presentation speed of the corrected video in the current study was extremely high with 52% of the corrected version being faster than 16 CPS and 32% faster than 20 CPS, as shown in Table 3.1. That is significantly above the threshold for comfortable reading with too many subtitles at a speed that would not be possible to read fully.

	Automated subtitles	Corrected subtitles
Total subtitles	422	533
Total words	3680	3735
Total characters (with spaces)	19675	20902
Duration (seconds)	1500	1489
Average presentation speed (number of characters/duration)	13	14.8
Subtitles between 15 and 19 CPS	98 (23%)	108 (20%)
Subtitles 16CPS and faster	142 (34%)	277 (52%)
Subtitles 20CPS and faster	44 (10%)	169 (32%)

Table 3.1. The analysis of the automated and corrected subtitles transcripts.

Another important aspect in relation to the automated subtitles is the rate of errors due to incorrect speech recognition. In fact, there are 982 errors in the automated subtitles, including 209 omissions but excluding a few remarks or questions by students that were not subtitled at all in the automated version. These omissions were included in the corrected version. The errors are caused due to misrecognition of the spoken words by the speech recognition software. The word accuracy was 73.3%, calculated by applying the formula used by the US National Institute of Standards and Technology (Dumouchel et al., 2011). Using the accuracy rate standard of 98% by the NER model (Romero-Fresco & Pérez, 2015) as a mere reference, the automated version has far too many errors to be useful. This is particularly true in an

academic context where accuracy of information is critical. It can therefore be expected that students would be frustrated by the high error rate and would either be distracted by the errors, or simply start ignoring the subtitles.

3.3.2 Questionnaires

A biographical survey about the participants' background and language history was collected. A pre-test with 10 multiple choice (MC) questions related to the video content was employed (Appendix A). This was used as baseline information indicating the amount of prior knowledge the participants have before viewing. A cognitive load test was used to determine the amount of self-perceived cognitive load while viewing subtitled or unsubtitled video (Appendix B). A post-test with 30 MC items was used to test the amount of content the students learned immediately after viewing (Appendix C).

3.3.2.1 Performance measurements

All the questions were taken from existing courses. The 10-items pre-test consisted of questions related to the 25 minutes video content. 5 MCs (questions 1-5) of the pre-test was taken from the MIT OpenCourseWare (2011) website under the topic Elasticity of Supply and Demand, the video used in the study. It is a topic included in the course on microeconomic principles. These 5 questions were constructed to test key vocabulary terms and understanding of key concept covered in the video. The other 5 MCs (questions 6-10) of the pre-test were drawn from the sample questions on the topic elasticity by Ralph R. Frasca (Frasca, 2007), the professor of Economics at the University of Dayton. The elasticity sample questions consisted of a total of 130 MC questions. Not all the 130 items are related to the concepts covered in the video, and those selected were screened against the video content to ensure item relevance. The 30-items post-test included the 10 items from the pre-test, with the addition of an extra 20 items from the elasticity sample questions by Frasca (2007). Those extra 20 items were also selected according to content covered in the video used in the experiment.

3.3.2.2 Cognitive load measurement

The cognitive load measurement is an adaptation from Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014). Usually a 1-item cognitive measurement is used to evaluate the cognitive load of a learner on a particular task. According to Paas et al. (2003), studies have shown that rating scales are fairly accurate and reliable based on the assumption that people are capable of introspecting their cognitive processes and report numerically on how much cognitive load they have expended. Leppink et al. (2014) developed a more precise instrument in measuring cognitive load which differentiates 3 types of cognitive loads, including intrinsic load, extraneous load and germane load. The instrument is a 13 items self-evaluated test, with items 1-4 measuring intrinsic load, items 5-8 measuring extraneous load, and items 9-13 measuring germane load.

The 30-item multiple choice questionnaire has an item reliability index of .986 and the 13-item cognitive load test has an item reliability index of .840. Reliability was calculated by means of Cronbach's Alpha (Table 3.2a and b).

Reliability Statistics on performance test

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.986	.990	30

(a)

Reliability Statistics on effort scale

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.840	.835	13

(b)

Table 3.2a. The reliability of the 30-item performance test. Table 3.2b. The reliability of the 13-item effort scale.

3.4 Design

This study is a quantitative experimental model using a 3 group (two test groups, one control group) test design and a pre-test-post-test design. Each group of students was randomly

assigned to one of the following conditions: English video with no subtitles (E), English video with English subtitles generated automatically (EA) and English video with corrected English subtitles (EC). Performance and the cognitive load of the participants were compared between the 3 tested conditions. By comparing the 2 subtitle conditions against the unsubtitled one, which is the control group, it helps to find answers to the three research questions by showing whether the intervention of subtitling has any difference in impact on performance and cognitive load. Participants were asked to perform a pre-test and a post-test before and after viewing the video. The pre-test was used as a baseline but also to measure the amount of prior knowledge the participants had on the topic of elasticity. The post-test was to determine the impact of subtitles, both EA and EC, on performance when comparing to the pre-test. Scores were also compared between EA and EC to determine the difference in impact if there is any. The cognitive load test was conducted to determine the self-perceived effort in viewing with or without subtitles.

3.4.1 Procedures

The experiment was conducted by 2 investigators in the scheduled tutorial sessions of the economics course. There were 7 groups in total. Each group of students was randomly assigned to one of the 3 conditions: condition 1 – English video with no subtitles (E); condition 2 – English video with English subtitles generated automatically (EA), and condition 3 – English video with corrected English subtitles (EC). A biographical survey and a pre-test were conducted before participants viewed the video. The video was about 25 minutes in length and was projected on a big screen for the participants to watch in one of the 3 conditions. All the conditions were identical for each of the groups including good sound and image. Two tests, in the order of 1) a cognitive load test and 2) performance post-test were given to the participants for completion immediately after the viewing. Data was collected, compared and analyzed.

3.4.2 Data analysis

The data was sufficiently normally distributed to allow for the use of an ANOVA. A one-way ANOVA was performed with dependent variables: full test performance, post-test performance, difference between pre- and post-test performance, intrinsic load, extraneous load or germane load) to determine whether there is any difference between the 3 conditions. Post hoc Tukey tests and reliability tests were carried out as well. A significance level of $\alpha = 0.05$ was adopted for all statistical analyses reported.

Chapter 4. Results

4.1 Introduction

The main aim of the study is to determine whether different quality of subtitles has an impact on performance and cognitive load. As discussed in the previous chapter, a performance test as well as a cognitive load test (both post-hoc) were used to determine the impact of the different modes on performance and load.

4.2 Reliability

Table 4.1a shows the reliability of the 30-item performance test was very high, with a Cronbach's Alpha of .986. Likewise, the reliability of the 13-item effort scale was acceptable with a Cronbach's Alpha of .840, as shown in Table 4.1b.

Reliability Statistics on performance test

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.986	.990	30

(a)

Reliability Statistics on effort scale

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.840	.835	13

(b)

Table 4.1 Reliability of instruments

4.3 Performance

A one-way ANOVA on the pre-test results showed no significant difference between the three groups in terms of prior knowledge on the topic, as shown in Table 4.2.

ANOVA

Pre-test percentage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	493.648	2	246.824	.798	.453
Within Groups	27527.004	89	309.292		
Total	28020.652	91			

Table 4.2. A one-way ANOVA on the pre-test results.

In terms of the performance on the full 30-item test, the group that saw unsubtitled video did slightly better than the group that saw the automatic subtitles, and the group that saw the corrected subtitles performed worst (see Figure 4.1 and Tables 4.3).

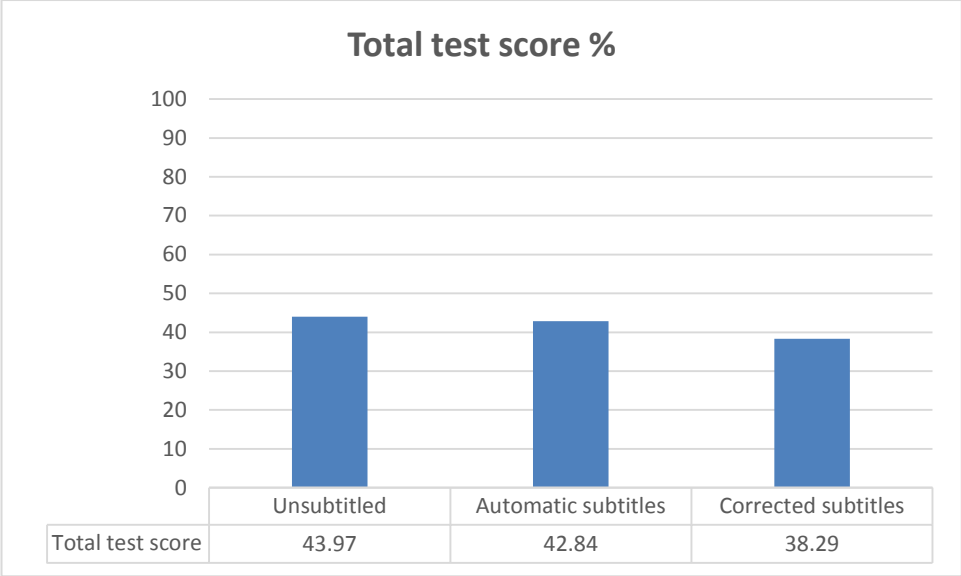


Figure 4.1. Full test performance

		N	Mean	Std. Deviation	Std. Error
Full test percent	Unsubtitled	21	43.968095240000004	9.346909446000000	2.039662860000000
	Automatic subtitles	34	42.843235290000000	13.685857710000000	2.347105234000000
	Corrected subtitles	37	38.288918920000000	15.307974830000001	2.516615559000000
	Total	92	41.268369570000004	13.639853640000000	1.422053044000000

Table 4.3. Distribution of the full test scores

However, these differences did not reach significance in a one-way ANOVA as can be seen in Table 4.4.

ANOVA of performance

		Sum of Squares	df	Mean Square	F	Sig.
Full test percent	Between Groups	565.839	2	282.920	1.539	.220
	Within Groups	16364.311	89	183.869		
	Total	16930.150	91			
	Within Groups	31551.234	89	354.508		
	Total	32347.826	91			

Table 4.4. A one-way ANOVA on the full test results

In terms of the difference between the performance on the 10-item pre- and post-test, it appears that the group that saw the video with corrected subtitles did not improve, and in fact did slightly worse as can be seen in Figure 4.2 and Table 4.5.

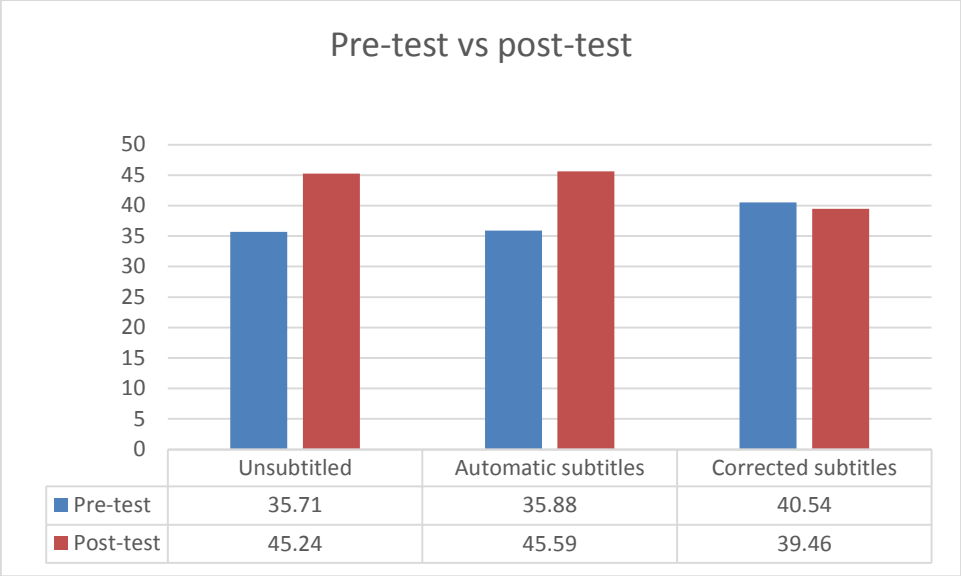


Figure 4.2. Comparison of performance on the 10-item pre- and post-test scores

		N	Mean	Std. Deviation	Std. Error
Difference between Pre- and Post-test	Unsubtitled	21	9.52	21.325	4.654
	Automatic subtitles	34	9.71	18.987	3.256
	Corrected subtitles	37	-1.08	22.458	3.692
	Total	92	5.33	21.404	2.232

Table 4.5. Distribution of the difference between pre- and post- test scores

The result indicates that students performed worse in the post-test after watching video with corrected subtitles, and with similar performance level in both unsubtitled and automated subtitles conditions. Figure 4.3 shows that the data of the pre- and post-test difference is normally distributed. The post-test used here to find out the difference is only the subset of the total post-video questionnaire, and is the same 10 items used in the pre-test. A one-way between subjects ANOVA was conducted to compare the effect of subtitles on performance in no subtitles, automated subtitles, and corrected subtitles conditions, and to determine whether

the difference between the pre-test and the post-test is significant. The difference between the pre- and post-test in fact approached significance with F(2,89) = 2.89, p = .061, as shown in Table 4.6.

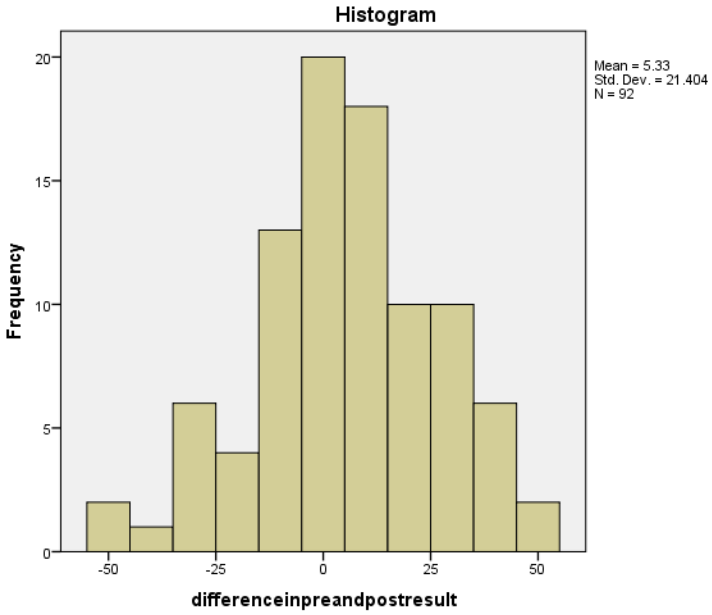


Figure 4.3. The mean score for pre- and post-test difference is normally distributed.

ANOVA of performance

		Sum of Squares	df	Mean Square	F	Sig.
Difference between Pre- and Post-test	Between Groups	2541.164	2	1270.582	2.888	.061
	Within Groups	39149.054	89	439.877		
	Total	41690.217	91			

Table 4.6. ANOVA results on difference between pre-and post-test

Post hoc comparisons using the Tukey HSD test indicated that the difference between automated and corrected subtitles conditions (p = 0.083) is closer to the significant level (p < .05), as shown in Table 4.7. However, the difference between unsubtitled and automated subtitles conditions (p = 0.999), and unsubtitled and corrected subtitles conditions (p = 0.159) are not significant. The Tukey test confirmed that the difference lies not between the unsubtitled condition and the subtitled conditions, but between the two subtitled conditions.

Multiple Comparisons

Dependent Variable: Difference between Pre- and Post-test

Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Unsubtitled	Automatic subtitles	-.182	5.821	.999	-14.06	13.69
	Corrected subtitles	10.605	5.730	.159	-3.05	24.26
Automatic subtitles	Unsubtitled	.182	5.821	.999	-13.69	14.06
	Corrected subtitles	10.787	4.983	.083	-1.09	22.66
Corrected subtitles	Unsubtitled	-10.605	5.730	.159	-24.26	3.05
	Automatic subtitles	-10.787	4.983	.083	-22.66	1.09

Table 4.7. Tukey test

4.4 Cognitive load

In terms of the cognitive load induced by the different modes, there was little difference as can be seen in Figure 4.4 and Table 4.8.

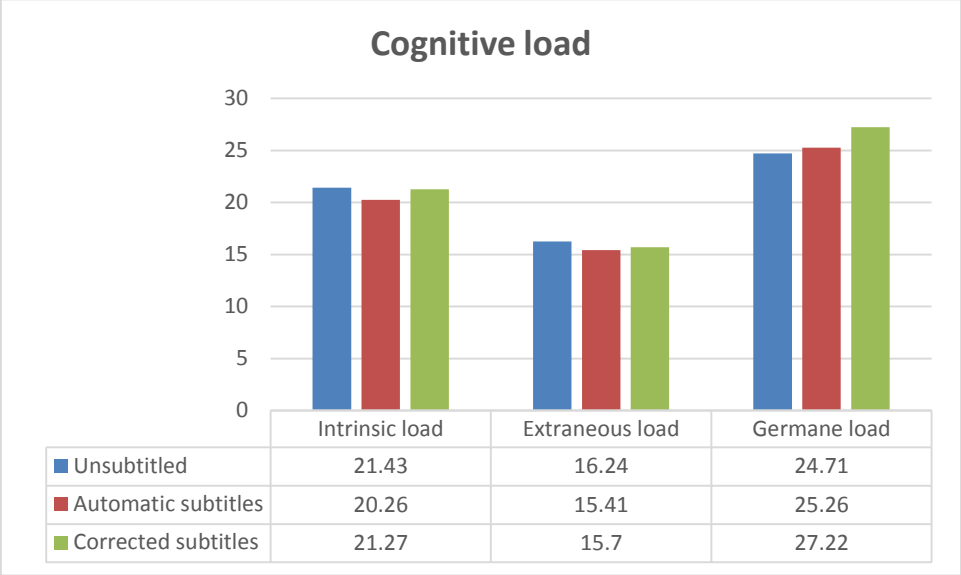


Figure 4.4. Comparison of the 3 types of cognitive loads on the 3 conditions

		N	Mean	Std. Deviation	Std. Error
Intrinsic load	Unsubtitled	21	21.43	4.434	.967
	Automatic subtitles	34	20.26	5.089	.873
	Corrected subtitles	37	21.27	5.881	.967
	Total	92	20.93	5.260	.548
Extraneous load	Unsubtitled	21	16.24	6.252	1.364
	Automatic subtitles	34	15.41	6.150	1.055
	Corrected subtitles	37	15.70	6.004	.987
	Total	92	15.72	6.055	.631
Germane load	Unsubtitled	21	24.71	6.958	1.518
	Automatic subtitles	34	25.26	4.857	.833
	Corrected subtitles	37	27.22	5.940	.977
	Total	92	25.92	5.860	.611

Table 4.8. The mean score of the cognitive load measurement

A one-way between subjects ANOVA was conducted to compare the impact of subtitles on 3 types of cognitive loads in no subtitles, automated subtitles, and corrected subtitles conditions, and to determine whether the difference in intrinsic load, extraneous load and germane load are significant. The result shows no significant effect of subtitles on cognitive load at the p< .05 level for the three conditions, as shown in Table 4.9.

ANOVA of cognitive load

		Sum of Squares	df	Mean Square	F	Sig
Intrinsic load	Between Groups	24.551	2	12.275	.438	.647
	Within Groups	2493.058	89	28.012		
	Total	2517.609	91			
Extraneous load	Between Groups	8.878	2	4.439	.119	.888
	Within Groups	3327.775	89	37.391		
	Total	3336.652	91			
Germane load	Between Groups	107.294	2	53.647	1.582	.211
	Within Groups	3017.174	89	33.901		
	Total	3124.467	91			

Table 4.9. A one-way ANOVA shows no significant effect of subtitles on cognitive load, p > .05

In sum, the study did not show any significant difference in either performance or cognitive load between the three modes, with the only difference approaching significance being in terms of the difference between pre- and post-test.

4.5 Discussion

The results in this study suggest that these readily-available subtitles (automated or corrected) have no significant effect on performance and cognitive load, although the corrected subtitles seemed to have a negative impact when compared with the automatic subtitles. Since these findings are counter to the body of literature suggesting that subtitles are beneficial to comprehension, we have to explore possible explanations. As indicated in Chapter 3, there were two significant characteristics of these two sets of subtitles that could have impacted on the performance, namely the low accuracy rate of the automatic subtitles, and the high presentation speed of the corrected subtitles. It should, for example, be noted that the presentation speed of the automated and corrected subtitles is very fast in general, especially in the corrected subtitles in with more than half of them being faster than 16 CPS. The data indicates that students performed worse when watching video with corrected subtitles. The lower mean score of the extraneous load indicates that subtitle presentation assists in information processing when comparing with intrinsic and germane load. The fact that students put more effort in processing the content in the video with no subtitles condition gives a general indication that subtitles are still beneficial when providing a favorable learning environment although none of these differences reached significance which makes it impossible to generalise from these findings.

The results of this study do not show that adding automatically generated or corrected subtitles to an educational video improve learning. The reason there was no significant difference could be ascribed to the fact that the error rate in the AE condition is substantially below respeaking standards. This would render the subtitles a distraction at best, and probably resulted in students ignoring the subtitles, something that will have to be verified with eye

tracking studies. The high presentation speed in the CE condition would make it virtually impossible for students to read around one third of the subtitles, which would have been a serious distraction and could easily interfere with comprehension. This seems to be supported by the findings.

In relation to the hypotheses, the results do not support the three hypotheses of the study that 1) automatically generated and corrected subtitles would result in better performance when compared to unsubtitled video; 2) automatically generated and corrected subtitles would result in lower cognitive load when compared to unsubtitled video. However, the trend of the data did show both AE and CE conditions result in lower intrinsic and extraneous load, and as a result meant that students would theoretically have more germane load available although they could not benefit from this in view of the problems with the two subtitled versions. Also, the lack of statistical significance means that this is at best a trend that will have to be established with refined experiments. In terms of Hypothesis 3), the corrected subtitles did not outperform the other 2 test conditions, instead the trend of the results indicated that students performed worst in the corrected subtitles, again the result is not statistically significant.

The trend of the current results shows that the presentation of subtitles assists information processing with less mental effort, in line with the results from Kruger et al. (2013) that subtitles reduced cognitive load of students in processing learning materials. There are a number of explanations for the presented findings. The major reason that this study does not have any significant results is that the presentation speed was so variable and so high in both automatically generated and corrected subtitles versions that it negated any possible benefits. The fact that 52% of the subtitles in the corrected version are 16 CPS and faster and 32% of its subtitles are 20 CPS and faster, made the corrected subtitles impossible to read. Even though the corrected version is very accurate, since the subtitles are created verbatim by ASR, the number of characters per second displayed to the reader is far too high. As concluded in the study of Romero-Fresco (Forthcoming) , fully verbatim subtitles are not desirable because

high presentation speed is caused mainly by the high speech rate of the speakers in the video, and it is far too fast to be read and comprehend despite its high accuracy rate.

The automated version is slightly slower than the corrected version, but the fact that it contains a high number of errors (there are 982 errors out of a total of 3680 words including omissions), negated the possible benefits of subtitle reading. The accuracy rate of the automated subtitles is 73.37% as calculated in the material section. When closely investigating the errors in the automated transcript, it appears that these errors were mainly caused by word misrecognition. The following are some of examples from the spoken word misrecognition: *empirical economist* was misrecognised as *pair click on a mess*; *correlation* as *poorly shop*; *useless* as *useful*; *children* as *troll*; *shift* as *ship*; *E1* as *you want* and *E2* as *he to*. The high error rate made the automated condition similar to the unsubtitled condition, and that the high presentation rate resulted in lower performance, thus also made the corrected condition similar to the unsubtitled condition. As indicated by the post hoc Tukey test, the difference shown in this study lies between the two subtitled conditions, not lies between the unsubtitled condition and the 2 subtitled conditions. However, the corrected subtitles did result in higher germane load which shows promise for the mode.

Chapter 5. Conclusion

5.1 Introduction

This study was conducted to investigate whether automatically generated or corrected subtitles would have an impact on performance and cognitive load. The literature suggests that subtitles would be beneficial in learning in terms of aspects such as vocabulary recognition and acquisition, retaining information, comprehension, and language acquisition in an educational context. Results show that the simultaneous presentation of both visual and auditory information actually assists learning due to the dual coding effect. Further investigations have been done in order to understand how subtitle reading would impact information processing and cognitive load. Three types of cognitive load are differentiated according to CLT. They are intrinsic load, extraneous load, and germane load. Each of these cognitive loads represents different loading in how the brain perceives an intended task. Intrinsic load is an interaction between the difficulty of the intended material and the learner's expertise level. Extraneous load is caused by the way the material is presented. Germane load is the remaining accessible cognitive resources people use to process information and automate schemas.

A few methodologies, such as subjective rating scale techniques, physiological techniques, and task-performance-based techniques, have been used regularly to evaluate cognitive load. Contradicting results have been found regarding whether subtitle reading would cause cognitive overload and thus decrease learning, and these contradictions may be caused by different cognitive measurements and study procedure. However, despite the contradicting results, the benefits of subtitles were consistently proven in the literature.

5.2 Research motivation

The increased use of online video lectures by large numbers of educational entities leads to a huge demand for large quantities of subtitled video. Automatically generated subtitles could

be a solution in meeting the need to produce subtitles fast and economically. Automated subtitles are created through speech recognition software, but usually with a high error rate that could potentially affect subtitle reading. It is only logical to correct these errors to improve readability. Since there are hardly any empirical studies that compare different types and quality of subtitles, it was the goal of this study to fill the gap of this research area.

5.3 Research findings

This study was conducted to investigate automatic and corrected subtitles as they are the types of subtitles that are normally available online. However these modes of subtitles do not seem to be beneficial under conditions of high error rates or high presentation speed. It would therefore seem that for subtitles to be beneficial, they would have to be corrected as well as presented at a reasonable reading speed. The real challenge will be to find ways of increasing the accuracy of transcripts, and even more so, of reducing the text automatically to bring down the presentation speed, and this means that the powerful mode of subtitling currently remains out of reach of the majority of institutions.

5.4 Implication

There are two possible implications to the issue on both subtitles being either too fast to process and comprehend, or having too many errors for comprehension. First, in order to have a reasonable presentation speed, this mode of subtitles can only be useful after being edited which includes summarising, reducing and reformulating. Professional subtitlers are still essential in editing the transcripts of the original text to meet subtitling standards (Diaz-Cintas & Remael, 2007). However, this implies that professionals would still be involved, resulting in the problem of costliness and time may be lessened but it is still an issue when large quantities of subtitles are required in a short time. The second implication is to control the audio visual recording environment of the video to the extent that the misrecognition by the speech recognition software would decrease. However there are a lot of variables in

controlling such an environment. Technology may advance to a point that it may be possible to have a highly sophisticated system that not only can recognise speech automatically but accurately, and also can perform simple reduction and reformulation of the original transcripts. Future research on technology may allow production of highly accurate subtitles automatically that is possible for readers to process and comprehend.

5.5 Limitation and further research

There are a few limitations to this study. The current study used an availability sample from year one students, and a larger sample size is essential in future research to yield possible significant results. The fact that this study used a video with short duration (25 minutes) and a single viewing, and with a nearly significant result implies that by manipulating the experimental environment differently in future studies, such as using a video with longer duration for a longer period of time in a longitudinal study, may possibly have a better research outcome. In future research, corrected subtitles at lower presentation speed and over a longer period of time such as a full school term will be required, as results would probably only start manifesting after more exposure with students becoming used to the mode. Furthermore, eye tracking studies should be done to determine to what extent the students actually try to read the subtitles.

References

Armstrong, A. W., Idriss, N. Z., & Kim, R. H. (2011). Effects of video-based, online education on behavioral and knowledge outcomes in sunscreen use: A randomized controlled trial. *Patient Education and Counseling, 83*(2), 273-277. doi:http://dx.doi.org/10.1016/j.pec.2010.04.033

Australian Government Department of Education and Training. (2016). *End of year summary international students enrolment data-Australia-2015* Retrieved August 30, 2016, from https://internationaleducation.gov.au/research/International-Student-Data/Documents/Monthly%20summaries%20of%20international%20student%20enrolment%20data%202015/12_December_2015_MonthlySummary.pdf

Berdichevsky, D. (2014, May 22). *What's the Difference Between YouTube's Automatic Captions, DIY Captions, and CaptionsForYouTube Captions?* : Message posted to http://www.captionsforyoutube.com/blog/posts/11-what's-the-difference-between-youtube's-automatic-captions,-diy-captions,-and-captionsforyoutube-captions?

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

Blake, J. (2015, Feberary 4). YouTube: We know automatic subtitles aren't good enough. *Newsbeat*. Retrieved from http://www.bbc.co.uk/newsbeat/article/31004497/youtube-we-know-automatic-subtitles-arent-good-enough

Bond, L. (2014a, November 24). *Are automatically generated captions and transcripts detrimental to video SEO?* : Message posted to

http://www.3playmedia.com/2014/11/24/are-automatically-generated-captions-transcripts-detrimental-video-seo/.

Bond, L. (2014b, April 4). *Wondering What YouTube Has to Say About Captioning YouTube Videos?* : Message posted to http://www.captionsforyoutube.com/blog/posts/2-wondering-what-youtube-has-to-say-about-captioning-youtube-videos.

Coursera. (2014, April 27). *Introducing Coursera's New Global Translator Community*: Message posted to http://coursera.tumblr.com/post/84088014661/introducing-courseras-new-global-translator.

Coursera. (n.d.). *Video Translation* Retrieved October 2 2016, from https://learner.coursera.help/hc/en-us/articles/208279836-Video-translations

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

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

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.

Diaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation : subtitling / Diaz-Cintas, Jorge & Aline Remael*. Kinderhook, NY: St. Jerome Pub.

Dumouchel, P., Boulianne, G., & Brousseau, J. (2011). Measures for quality of closed captioning. In A. Şerban, A. Matamala, & J.-M. Lavaur (Eds.), *Audiovisual Translation in Close-up: Practical and Theoretical Approaches* (pp. 161-172). Bern: Peter Lang.

Frasca, R. R. (2007). *Chapter 4 - Elasticity - Sample Questions* [PDF file]. Retrieved from http://academic.udayton.edu/pmic/mc%20questions/chap%204%20mc.pdf

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

Harrenstien, K. (2009, November 19). *Automatic captions in YouTube*: Message posted to https://googleblog.blogspot.com.au/2009/11/automatic-captions-in-youtube.html.

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

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition / Daniel Jurafsky, James H. Martin* (2nd ed. ed.). Upper Saddle River, N.J.: Upper Saddle River, N.J. : Pearson Prentice Hall.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1.

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. *The Cambridge Handbook of Multimedia Learning*: Cambridge University Press.

Khan Academy. (n.d.). *How are the English captions created?* Retrieved October 2 2016, from https://khanacademy.zendesk.com/hc/en-us/articles/226430308-How-are-the-English-captions-created-

Khesin, T. (2012, October 2). *MIT OpenCourseWare launches interactive transcripts and video search*: Message posted to http://www.3playmedia.com/2012/10/02/mit-opencourseware-launches-interactive-transcripts-video-search/.

Kruger, J. L. (2013). Subtitles in the classroom: balancing the benefits of dual coding with the cost of increased cognitive load. *Journal for Language Teaching, 47*(1), 29-54. doi:10.4314/jlt.v47i1.2

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

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.

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

Lockrey, M. (2015, July 25). *YouTube automatic craptions score an incredible 95% accuracy rate!* : Message posted to https://medium.com/@mlockrey/youtube-s-incredible-95-accuracy-rate-on-auto-generated-captions-b059924765d5#.femfbg6cs.

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

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.

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and instruction, 21*(6), 687-704. doi:http://dx.doi.org/10.1016/j.learninstruc.2011.03.004

MIT OpenCourseWare. (2010, November 22). *Help MIT subtitle OpenCourseWare videos*: Message posted to https://about.amara.org/2010/11/22/help-mit-subtitle-opencourseware-videos/.

MIT OpenCourseWare. (2011). Principles of microeconomics: elasticity of supply and demand. Retrieved June 15, 2016, from http://ocw.mit.edu/courses/economics/14-01sc-principles-of-microeconomics-fall-2011/unit-1-supply-and-demand/elasticity/

MIT OpenCourseWare. (2012, January 24). *Lecture 3 - Elasticity* [Video file]. Retrieved from https://www.youtube.com/watch?v=Ye4vL7u6N2g

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

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. (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

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

Romero-Fresco, P. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics, 6,* 109-133. Retrieved December 12, 2016, from http://www.mapaccess.org/publications/journal-article/more-haste-less-speed-edited-versus-verbatim-respoken-subtitles

Romero-Fresco, P. (Forthcoming). *Quality of live caption: A Qualitative and Quantitative Assessment*.

Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy Rate in Live Subtitling: The NER Model. In R. B. Pineiro & J. Diaz-Cintas (Eds.), *Audiovisual Translation in a Global Context. Mapping an everchanging landscape* (pp. 28-50). London: Palgrave Macmillan.

University of Washington. (2016). *Can captions be generated automatically using speech recognition?* Retrieved August 4, 2016, from http://www.washington.edu/accesscomputing/can-captions-be-generated-automatically-using-speech-recognition

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

Wald, M., & Bain, K. (2008). Universal access to communication and learning: the role of automatic speech recognition. *Universal Access in the Information Society, 6*(4), 435-447. doi:10.1007/s10209-007-0093-9

Wilson, E. A. H., Park, D. C., Curtis, L. M., Cameron, K. A., Clayman, M. L., Makoul, G., vom Eigen, K., & Wolf, M. S. (2010). Media and memory: The efficacy of video and print materials for promoting patient education about asthma. *Patient Education and Counseling, 80*(3), 393-398. doi:http://dx.doi.org/10.1016/j.pec.2010.07.011

YouTube. (n.d.). *Add subtitles & closed captions* Retrieved October 2 2016, from https://support.google.com/youtube/answer/2734796?hl=en&ref_topic=3014331

Zhang, Y., & Mi, Y. (2010). Another Look at the Language Difficulties of International Students. *Journal of Studies in International Education, 14*(4), 371-388. doi:10.1177/1028315309336031

Appendix A

10-item multiple choice questions pre-test (Frasca, 2007; MIT OpenCourseWare, 2011)

Elasticity Quiz

1) Which of the following accurately characterize perfectly inelastic demand?

a. The demand curve is vertical but does not change regardless of what happens to price.

b. The demand curve is vertical.

c. The demand curve is horizontal.

d. Demand does not change regardless of what happens to price.

2) When do we expect to see a perfectly elastic demand curve?

a. When a good has many complements.

b. When a good has a perfect substitute.

c. When a good has no substitutes.

d. When there is limited supply of a good.

3) Let's say a researcher makes a study of patients in hospitals and finds they are much sicker than the average person in the population. Then he concludes that hospitals make patients sick. The researcher is mixing up two concepts; what are they?

a. Demand and supply.

b. Correlation and causation.

c. Theory and empirics.

d. Shifts along the demand curve and shifts in the demand curve.

4) If the elasticity of demand for a good is sufficiently negative, firms may actually lose revenues when they raise the price of the good. Why is this?

a. The elasticity of demand changes.

b. Consumers substitute from other goods to buy this firm's good.

c. The supply curve shifts in.

d. Fewer people buy the good at the higher price, and so overall revenues are lower.

5) What is an example of a supply shock in the orange market that would enable us to estimate demand elasticity?

a. All of these.

b. A plant disease that hits orange crops.

c. A new government tax on orange growers.

d. An early frost in Florida that destroys crops.

6. When the quantity of coal supplied is measured in kilograms instead of pounds, the demand for coal becomes

a. more elastic.

b. neither more nor less elastic.

c. less elastic.

d. undefined.

7. An increase in subway fares in New York City will boost your expenditures on subway rides if

a. the supply of subway rides is elastic.

b. the supply of subway rides is inelastic.

c. your demand for subway rides is inelastic.

d. your demand for subway rides is elastic.

8. The demand for Honda Accords is

a. probably inelastic and less elastic than the demand for automobiles.

b. probably elastic but less elastic than the demand for automobiles.

c. probably elastic and more elastic than the demand for automobiles.

d. probably inelastic but more elastic than the demand for automobiles.

9. Which of the following is likely to have the smallest price elasticity of demand?

a. a new Ford automobile

b. a new automobile

c. a new Ford Mustang

d. an automobile

10. Demand is inelastic if

a. a leftward shift of the supply curve raises the total revenue.

b. the good in question has close substitutes.

c. the smaller angle between the vertical axis and the demand curve is less than 45 degrees.

d. large shifts of the supply curve lead to only small changes in price.

Appendix B

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

Please respond to each of the questions by circling the most applicable number on the following

scale 0-10 (0 = not at all the case and 10 = completely the case).

1) The content covered in the video was very complex.

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

2) The video covered formulas that I perceived as very complex.

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

3) The video covered concepts and definitions that I perceived as very complex.

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 the complexity of this video

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

5) The explanations in the video were very unclear.

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

6) The explanations were full of unclear language.

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

7) The explanations 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

8) I invested a very high mental effort in unclear and ineffective explanations in this video.

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

9) The video really enhanced my understanding of the content that was covered.

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

10) The video really enhanced my understanding of the formulas that were covered.

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

11) The video really enhanced my knowledge of concepts and definitions that were covered.

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

12) The video really enhanced my knowledge and understanding of the elasticity of supply and demand.

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

13) I invested a very high mental effort during this video in enhancing my knowledge and understanding.

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

Appendix C

30-item multiple choice questions post-test (Frasca, 2007; MIT OpenCourseWare, 2011)

Elasticity Quiz after viewing

1. The price elasticity of demand depends on

a. the units used to measure price but not the units used to measure quantity.

b. the units used to measure price and the units used to measure quantity.

c. the units used to measure quantity but not the units used to measure price.

d. neither the units used to measure price nor the units used to measure quantity.

2. The demand for food is most elastic in countries

a. with low income levels.

b. that are highly urbanized.

c. with intermediate income levels.

d. with high income levels.

3. Demand is perfectly inelastic when

a. the good in question has perfect substitutes.

b. shifts in the supply curve results in no change in price.

c. shifts of the supply curve results in no change in quantity demanded.

d. shifts of the supply curve results in no change in the total revenue from sales.

4. What is an example of a supply shock in the orange market that would enable us to estimate demand elasticity?

a. All of these.

b. A plant disease that hits orange crops.

c. A new government tax on orange growers.

d. An early frost in Florida that destroys crops.

5. Which of the following accurately characterize perfectly inelastic demand?

a. The demand curve is vertical but does not change regardless of what happens to price.

b. The demand curve is vertical.

c. The demand curve is horizontal.

d. Demand does not change regardless of what happens to price.

6. The demand for Honda Accords is

a. probably inelastic and less elastic than the demand for automobiles.

b. probably elastic but less elastic than the demand for automobiles.

c. probably elastic and more elastic than the demand for automobiles.

d. probably inelastic but more elastic than the demand for automobiles.

7. If the elasticity of demand for a good is sufficiently negative, firms may actually lose revenues when they raise the price of the good. Why is this?

a. The elasticity of demand changes.

b. Consumers substitute from other goods to buy this firm's good.

c. The supply curve shifts in.

d. Fewer people buy the good at the higher price, and so overall revenues are lower.

8. The slope of a demand curve depends on

a. the units used to measure quantity but not the units used to measure price.

b. the units used to measure price and the units used to measure quantity.

c. the units used to measure price but not the units used to measure quantity.

d. neither the units used to measure price nor the units used to measure quantity.

9. Producers' total revenue will decrease if

a. the price rises and demand is inelastic.

b. income increases and the good is a normal good.

c. the price rises and demand is elastic.

d. income falls and the good is an inferior good.

10. A good with a vertical demand curve has a demand with

a. infinite elasticity.

b. unit elasticity.

c. zero elasticity.

d. varying elasticity.

11. Which of the following is likely to have the smallest price elasticity of demand?

a. a new Ford automobile

b. a new automobile

c. a new Ford Mustang

d. an automobile

12. A good with a horizontal demand curve has a demand

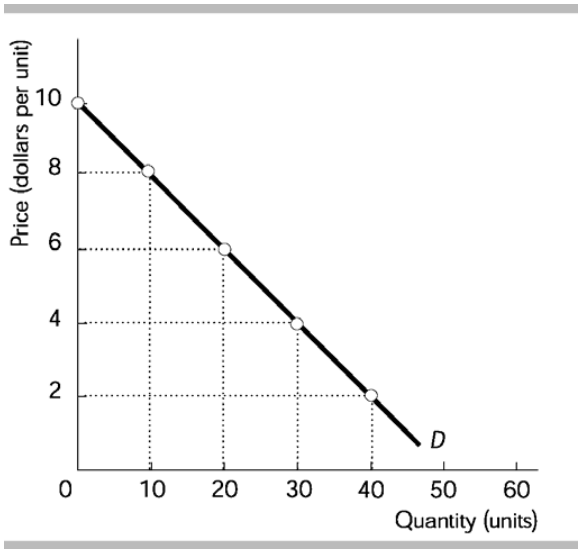
a. with an income elasticity of demand of 0.

b. with a price elasticity of demand of infinity.

d. for which there are no substitute.

d. with a price elasticity of demand of 0.

Refer to the following diagram to answer 13-15.



13. The figure above illustrates a linear demand curve. If the price falls from $8 to $6,

a. the quantity demanded will increase by less than 20 percent.

b. total revenue will remain unchanged.

c. total revenue will increase.

d. total revenue will decrease.

14. The figure above illustrates a linear demand curve. In the range from $8 to $6,

a. the demand is unit elastic.

b. the demand is price inelastic.

c. the demand is price elastic.

d. more information is needed to determine if the demand is price elastic, unit elastic, or inelastic.

15. The figure above illustrates a linear demand curve. If the price falls from $6 to $4,

a. total revenue will decrease.

b. total revenue will increase.

c. quantity demanded will increase by more than 100 percent.

d. total revenue will remain unchanged.

16. Let's say a researcher makes a study of patients in hospitals and finds they are much sicker than the average person in the population. Then he concludes that hospitals make patients sick. The researcher is mixing up two concepts; what are they?

a. Demand and supply.

b. Correlation and causation.

c. Theory and empirics.

d. Shifts along the demand curve and shifts in the demand curve.

17. The more substitutes available for a product,

a. the larger is its income elasticity of demand.

b. the smaller is its income elasticity of demand.

c. the smaller is its price elasticity of demand.

d. the larger is its the price elasticity of demand.

18. Of the following, demand is likely to be the least elastic for

a. Toyota automobiles.

b. compact disc players.

c. Ford automobiles.

d. toothpicks.

19. Of the following, demand is likely to be the least elastic for

a. pink grapefruit.

b. iceberg lettuce.

c. insulin for diabetics.

d. diamonds.

20. If a rise in the price of good B increases the quantity demanded of good A,

a. B is a substitute for A, but A is a complement to B.

b. A is a substitute for B, but B is a complement to A.

c. A and B are complements.

d. A and B are substitutes.

21. Supply is elastic if

a. a 1 percent change in price causes a larger percentage change in quantity supplied.

b. the good in question is a normal good.

c. the slope of the supply curve is positive.

d. a 1 percent change in price causes a smaller percentage change in quantity supplied

22. When do we expect to see a perfectly elastic demand curve?

a. When a good has many complements.

b. When a good has a perfect substitute.

c. When a good has no substitutes.

d. When there is limited supply of a good.

23. If at a given moment, no matter what the price, producers cannot change the quantity supplied, the momentary supply

a. has infinite elasticity.

b. has unit elasticity.

c. does not exist.

d. has zero elasticity.

24. The elasticity of supply measures the sensitivity of

a. supply to changes in costs.

b. quantity supplied to a change in price.

c. price to changes in supply.

d. quantity supplied to quantity demanded

25. An increase in subway fares in New York City will boost your expenditures on subway rides if

a. the supply of subway rides is elastic.

b. the supply of subway rides is inelastic.

c. your demand for subway rides is inelastic.

d. your demand for subway rides is elastic.

26. The demand for a good is elastic if

a. a decrease in its price results in a decrease in total revenue.

b. the good is a necessity.

c. an increase in its price results in an increase in total revenue.

d. an increase in its price results in a decrease in total revenue.

27. If a price decrease results in your expenditure on a good decreasing, your demand must be

a. unit.

b. inelastic.

c. linear.

d. elastic.

28. The route from Dallas to Mexico City is served by more than one airline. The demand for tickets from American Airlines for that route is probably

a. elastic and more elastic than the demand for all tickets for that route.

b. inelastic and less elastic than the demand for all tickets for that route.

c. elastic but less elastic than the demand for all tickets for that route.

d. inelastic but more elastic than the demand for all tickets for that route.

29. When the quantity of coal supplied is measured in kilograms instead of pounds, the demand for coal becomes

a. more elastic.

b. neither more nor less elastic.

c. less elastic.

d. undefined.

30. Demand is inelastic if

a. a leftward shift of the supply curve raises the total revenue.

b. the good in question has close substitutes.

c. the smaller angle between the vertical axis and the demand curve is less than 45 degrees.

d. large shifts of the supply curve lead to only small changes in price.