Investigating the impact of including videos or still images in computer-based academic listening comprehension tests

Iman Y Elmankush

Doctor of Philosophy

University of York

Education

April 2017

Abstract

Visual materials are central to second language listening (L2), yet their use in L2 listening assessment is very limited. Disagreement about the listening abilities that should be included in L2 listening construct and absence of empirical evidence about the effects of visual materials on performance in L2 listening tests led to disagreement about their use in L2 listening tests. Moreover, previous research did not explore how test takers interact with visual materials other than video texts. The present research attempts to contribute to the existing literature by exploring test takers' viewing patterns using eye tracking technology with both video and still photo texts during L2 academic listening test. In addition, cued retrospective reports are employed to extract test takers' perceptions about the two types of visual materials and shedding light on the underlying cognitive processes they employed. Mixed-research method based on triangulation design was used to investigate test takers' (n = 30) performance in video, still photo, and audio texts, in addition to recording their viewing patterns, and their reported perceptions about the visual materials. The results revealed that test takers' performance in the video texts was superior to both still photos and audio texts, with statistically significant difference to audio texts. Cued retrospective report data showed higher helpful perceptions by test takers related to video texts with strong correlation to their scores, while still photo texts were perceived with higher distractedness. Eye tracking data partially coincided with the rest of the results, with two out of three measures- Fixation counts and Total dwell time- found to be higher with video texts. Implications of the study are that visual materials, especially video texts, should be considered in L2 listening tests as they present better representations to the target language use domain, which requires reconsidering the current L2 listening construct.

Table of Contents

Page

List of Tables

List of Figures

Page

Acknowledgement

I express my deep gratitude to my professor Leah Roberts for her exceptional support during this journey, for her insightful comments and immediate helpful feedback on the drafts of my thesis, and for her valuable discussions that helped me shape the ideas in this study. I am also grateful to Dr Irena Kuzborsca for her valuable and productive thesis advisory meetings, to Kurt Debono for his great patience and support in designing and producing my eye tracking experiment, to Marcus Johnson for his assistance with eye tracking complications. I am also deeply thankful to Cylcia Bolibaugh, David Gundry, Eleanor McCullough, Nora McIntyre and Fatma Layas for their help and cooperation in recording video lectures for the test in this study. I would also thank Francis Duah for his help and guidance with my data analysis. Special thanks to all my friends who encouraged me throughout my journey, especially Ghazal Syed for her kind comments and advice. Lastly, I would like to thank my beloved parents, for their endless support throughout my life, my dear husband for his patience, help, advice, and for all that he has done for me.

Author’s Declaration

I declare that the work in this thesis is a presentation of original work, and I am the sole author. This work has not, in whole or part, previously been presented for an award at this, or any other, University. All sources are acknowledged as Referenced.

Chapter One Introduction

1.1. Background

Second language (L2) listening is recognized as a vital component of language competence and attention to this skill by language researchers has noticeably increased in recent decades (Vandergrift and Goh, 2012; Wagner, 2010) after an era of neglect (Vandergrift, 2004). Despite this, it is still one of the least understood skills and most difficult to study and assess (Vandergrift, 2015). In an age of globalization, with English as an international language and medium for communication, millions of students, and professionals from all over the world are seeking academic education and employment in different countries where English is either the dominant language or a second language and is used as lingua franca (Suvorov, 2013). This increasing demand for English language education has consequently raised the need for international standardized language tests. Thus, a number of standardized English language tests has evolved, and have initially been introduced in English speaking countries like the United Kingdom, the USA, and Australia and then spread around the globe (Gallagher, 2003). Examples of some of the most common high-stake standardized English tests are IELTS (the International English Language Testing system), PTE academic (Pearson Test of English 'academic'), and TOEFL (the Test of English as a Foreign Language) (Field, 2011; Suvorov, 2013).

In order to meet the demands of different test applicants and also to cope with the technological development in all academic domains, many leading language-testing companies have developed a specialised test administration system and moved from paper-based tests (PBT) to computer-based tests (CBT) (Geranpayeh & Taylor, 2013). As the attention to computer-based tests has increased over the past two decades, some researchers have attempted to highlight the advantages of such tests (Douglas and Hegelheimer, 2007). CBT plays a vital role in the field of language assessment, and that is mainly because it employs multimedia content forms like images, video, audio, or texts, which makes it more flexible and inviting for test makers to develop tests that resemble real-life situations. Bearing in mind the importance of visual information as a component of multimedia, Clark and Mayer (2011) cites that several instructors and experts in language tests have confirmed the benefits of the use of visuals in CBT. Most of these materials have realistically considered the same conditions for test-takers that are likely to be used in real situations where target language is used (Lynch, 2011).

This importance of visuals is explicitly described by some researchers as an international mode of discourse by which people can easily communicate in almost all knowledge domains, by providing them with materials that support and complement the verbal input like geographical maps and anatomy illustrations and images, or even by the acts of the speakers when they use their body language and face expressions during their speech (Rowley-Jolivet, 2002).

1.2. Statement of the problem

The idea of using visual materials in listening tests seems rational to almost any person, as this would make a perfect sense. You can hear and see the person who is talking, just like in real life. However, this is not the case yet in language tests. Even though several researchers, test-makers and educators have emphasized the benefits of using the visual component in CBT, its use in listening assessment remains limited (Field, 2008; Rost, 2013; Vandergrift, 2015). There are different reasons that can be linked to the general reluctance in using visuals, like technology for instance. In many cases, poor internet connectivity, the need for specialized IT technicians to operate and update certain software, or the complexity involved in developing quality visual resources made test-designers hesitant to implement visuals in language tests (Ockey, 2007; Suvorov, 2013). These issues can be sensed to a greater level in some parts of the world more than others, like the case in developing countries. Though many technological hitches that used to be experienced by educators and researchers in the past do not exist currently and solutions are being introduced every day to reduce the impact of these problems, educators and researchers still do not effectively employ visuals in the L2 listening test for different reasons:

One of the reasons is the complex nature of listening. The complexity of the construct often leads to divergence among scholars regarding the explicit capabilities that listening assessments should verify. Would the use of visuals reflect the abilities that should be involved in listening or it would add an extra burden on the test-takers' shoulders as it conveys additional information? While some researchers are for the incorporation of visuals, others such as Buck (2001) Batty (2015), Coniam (2002), and Yousofi, Davoodi & Razmeh, 2015) are strongly opposing to the idea of including visuals in the L2. According to the opponents, the listening aspect should not include the capability to comprehend and use the visible information. They argue that listening should only consider the language comprehension ability of the task takers. This would include but not limited to semantics, syntax, and phonology and non-linguistic knowledge such as world knowledge and. Taking this viewpoint has led some of those who are against the idea of implementing visuals into listening tests such as Bucks (2001) to argue that the concept of visuals will eventually change the listening assessments to tests that examine multiple language skills and not focusing on listening per se. Conversely, exponents argue that such views ignore the modern changes in the integrated language assessments, which recommend that construct definitions in language should consider a range of abilities and not to be restricted to one skill at a time and strip it from any other abilities that can be naturally linked to it in real-life encounters (Cumming, 2013; Gebril & Plakans, 2013; Suvorov, 2013; Wolfersberger, 2013). These abilities can range from comprehending the verbal-linguistic input to interpreting non-verbal information (e. g., body language) or any source materials (linguistic, non-linguistic, or visual) that are available in the context. The rationale for this kind of integration according to Plakans and Gebril (2013) is to simulate the type of language normally used in real academic settings, which can motivate students and improve their performance.

Another reason why visual materials are not commonly used in listening assessments is the confusion over the type of visuals that should be used. Visuals come in different shapes and via different formats such as video clips, graphs, still-pictures, etc., and each of these types has many other different categories within it. It would take an extensive period of time and effort and using many different resources in order to come up with a clear idea of which of the available types of visuals can be used and what impact they might have on the test-taker. However, despite the variety of visuals, there is a lack of research on the real impacts of these materials on the learners' performance. Some scholars have concluded that employing visual materials in listening assessments has impacted positively on the performance of the learners, which has further enhanced their spoken language (Ginther, 2002; Rubin, 1994; Secules, Herron, & Tomasello, 1992; Suvorov, 2013). On the other hand, other scholars do not find any relevant significance of visual materials on the learners’ performance (Coniam, 2002; Gruba, 1993. Therefore, it is not surprising that several text designers have avoided the use of visual materials in listening tests because of a lack of empirical evidence on their significance. Hence, more research is needed in order to determine the effects of visuals on students’ performance, so test developers can decide whether to implement such materials in future L2 listening tests.

As the need for more research on the effects of visual materials on test takers’ performance has stemmed largely from the inconclusive results of previous research; one should inspect the reasons that led to such a great difference among researchers on the use of visuals. One very important factor to which the different results about the use of visuals can be attributed is research designs. A closer look at the studies available in the literature gives a clear indication that most of these studies compare the effects of using visuals on test takers’ performance in listening comprehension tests to audio-only version of the same test (Coniam, 2002; Gruba, 1993; Rubin, 1994; Suvorov, 2009). While this type of traditional comparison seems logical at first; it does not take into consideration any other factors besides the test takers’ scores in visual tests compared to their scores in audio-only tests. In other words, these studies which focused mainly on the outcome (product) of listening, had attributed the difference in test takers’ performance to the visual materials only. Here, the main problem with this assumption is that it does not capture the whole process of how visual materials can affect performance. In fact, it cannot simply be assumed that the difference in test takers’ performance can be completely accredited to visual materials without investigating the process which led to the scores, like their perceptions and supporting that with their viewing behaviour. Examining viewing behaviour can reveal important facts about how test takers interact with visual materials. It provides information about whether participants in fact watch the visuals in the first place or not. In addition, it gives a clear map of the distribution of overt visual attention of participants in terms of the objects they are watching, for how long, and in what order (Scheiter and Van Gog; 2009). In addition, teaming the viewing behaviour with the test takers’ perceptions would shed more light on the bigger picture of taking a visualised listening test by explaining and justifying the reasons for their viewing behaviour and the way they process the spoken text and answer the test questions. All these factors might together give an implication of the underlying cognitive processes that test takers employed during the test, which can help the test designer to improve the design of L2 listening tests in order to reflect target real-life situations, which may help to trigger cognitive processes similar to the ones employed in

real-life.

With all these factors in mind, it is clear that the cognitive processing of test-takers can be affected by the smallest change in their test-taking experience. However, up to the present time there seems to be a missing link in the process of studying the effects of visual materials on learners' performance in L2 listening comprehension tests. The viewing behavior of test-takers is not examined close enough to reveal how exactly they utilize these materials. Suvorov (2013) attributes that to the assumption that any detected difference in the performance of the individuals being assessed in the tests that compare audio-only to visual materials are automatically attributed to visuals. Such assumptions cannot be generalized because they do not give a detailed analysis of the learners' behavior while taking the tests of who may have viewed the materials for longer durations or not long enough to have any significant alterations of their performance. However, research done by Ockey (2007), and Wagner (2006b) attempted to address this gap. In both studies, the researchers used some digital video cameras to record and view behaviors of learners being assessed while taking a listening test. Thereafter, the researchers used a stopwatch to determine the time the test takers had made eye contact with the screens.

This technique of examining test takers' viewing behaviour is not sufficient for exploring the exact patterns of their viewing to the screen, and it can only give a rough idea about how long they made eye contact with the visuals. More recently, Suvorov (2013) conducted an eye tracking study to examine test takers' viewing behaviour to two types of videos, namely context and content video. In his study, Suvorov used authentic materials which were academic lectures recorded and uploaded into their universities YouTube channels, where context videos presented a lecturer giving a lecture in an auditorium (showing the context where the lecture took place), and content videos presented a lecturer giving a lecture and is supported by pictures or some kind of illustrations to explain the content of the lecture. This study is recognised as the first study that used eye tracking technology in order to accurately measure test takers viewing behaviour during L2 listening tests. It focused on different types of video texts only. It did not compare it to the viewing behaviour of still images which are regarded as an important type of visual materials, and which are already in use in some standardised tests like TOEFL computer-based test. Also, Suvorov’s (2013) study did not attempt to investigate the underlying cognitive processes that are used by listeners during taking a visualised listening test in order to understand how these visual materials affect test takers’ performance. The current study attempts to address these limitations in Suvorov’s (2013) study and by that it intends to make a unique contribution to the available body of literature.

As noted above, examining the viewing behaviour of L2 listening comprehension test takers seems to be in its very early and exploratory stages. Listening tests enhanced by different types of visual materials still need to be examined further in order to find out how the use of these types of visuals can affect performance. Does one type of visual have an advantage that positively affects test takers’ performance over another? If so, how? This goal cannot be accomplished without comparing the test takers’ performance in the different visual conditions, in this study video and still photos, and associating this with verbal data in order to find out how test takers perceive the different types of visuals, the reasons for their perceptions, and how their performance is affected by each type. In addition, exploring the test takers’ viewing behaviour has the potential to reveal some important information about their viewing patterns, which can add a new dimension that helps to explain what aspects they watched in the visualised texts and for how long. This procedure may possibly give some indications about the test takers’ cognitive processes during watching the visualised test, and that can help to understand more specifically the role of visual materials on their performance.

1.3. Purpose of the study

It is clear from the review of literature that more research is still needed to validate the use of visual materials in listening comprehension tests. Particularly, there is currently a lack of research comparing the effects of using video and still photos. Therefore, the main objective of the current study is to assess the impact of using visual materials (i.e., Video and still photos) on the performance of test-takers in listening comprehension tests by examining the test takers’ viewing behaviour and analysing their individual perceptions regarding the use of visuals.

The study is based on a proficiency test of visualized academic L2 listening comprehension, which aims to provide a more accurate representation of the target language use domain (academic lectures), by replicating lecture materials with the use of visual materials. The use of visual materials is believed to provoke authentic cognitive processes that play when learners observe academic lectures. More precisely, this study is structured to accomplish three main goals:

1) The study attempts to examine the potential differences that might exist in test takers' performance affected by different audio-visual materials, operationalized as video, still photos, and audio-only texts in L2 listening comprehension tests.

2) By exploring the test takers' perceptions of the visual materials- by conducting a cued retrospective report- some insights about their inner cognitive processing of the visual information provided by two visual texts- video and still photos- in the L2 listening test, can be revealed.

3) The study is also concerned with examining the different viewing behaviours of test takers by means of eye tracking technology.

In sum, the study intends to combine two important lines of research, namely the process and the product of listening. The first part of the study is a product-oriented, as it merely investigates the test takers' scores in the three types of texts used in the test. The latter parts are process-oriented since their focus is on gaining a deeper and more accurate level of understanding of the cognitive processes associated with L2 listening comprehension. These three main purposes of the study will be accomplished by implementing a mixed method research design. This study will largely rely on Creswell and Plano Clark's (2011) mixed research method involving gathering quantitative data and thereafter collecting qualitative data. The quantitative data will be the listening test scores and the eye tracking numeric and behavioural data, and the qualitative data are verbal data collected from test takers via the cued retrospective report.

The primary objective of this research is to focus on the academic listening target language use. In order to achieve that it aims to produce materials that reflect academic listening, specifically academic lectures. The purpose of doing this is to ensure that the characteristics of the test tasks reflect the characteristics of this particular domain, and it can trigger performance similar to the one that happens in real-life situations and to capture the L2 listening skills and abilities required for academic listening situations. This is done to reduce any potential factors that could cause invalidity and to prevent the under-representation of the construct. Therefore, the first goal of the study will be accomplished by gathering performance data from the L2 academic listening test that has been created specifically for this study. This test is divided into three parts: 1) the traditional audio-only texts, 2) a video text, 3) still photos text. The study recognizes the listening construct employed in this test as the ability to comprehend and process information from acoustic and visual inputs in academic lectures within university settings where English is the primary language of communication.

The choice of this particular construct is based on the key purpose of academic listening tests, which is basically discriminating between test takers who seek a place in universities where English is the main means of communication. More precisely, because in our modern world almost all types of university lectures and academic situations involve some type of visual materials, the inclusion of *visual input* besides the *acoustic input* in the construct of listening is regarded as an absolute necessity if the test intends to reflect the target language use domain. However, the inclusion of the visual input besides the acoustic input in the test might lead to some ambiguity and puzzlement regarding naming the test Listening Test. In other words, it is not very clear to what extent would this term Listening Test still be suitable for a test supported with visuals, as in a way it does not purely test the test takers' listening ability, but it adds the visual aspect as well. The novelty of this branch of research makes it difficult to come up with a new term of this type of tests and because the test used in this study is using traditional listening-only texts besides the visual ones, it is decided to stick to the term Listening Test, while encouraging more efforts to improve the terminologies related to this type of tests in the future.

The second goal will be achieved by implementing cued retrospective reports (Van Gog, Paas, Van Merriënboer, and Witte, 2005). Cued retrospective reports are described by Van Gog and her colleagues as "process-tracing technique" (2005, p. 237), that can reveal information about actions and movements performed by the subjects, the reason of their performance, and in what way. This type of verbal report allows the researcher to collect verbal information from test takers immediately after finishing the test, by showing them a record of their eye movements, in the current study, in the form of heat maps. These maps allow test takers to give more accurate verbal descriptions of the way they processed the visual information in the texts and answered the test questions, and consequently, increase the potential to reveal some valuable insights into the cognitive process employed by the test takers during the test. The analysis of this type of data is very significant as it is believed to give a profound understanding of the way test takers have used the visual information in processing the texts and answering the test questions.

Eye tracking technology will be used to generate heat maps for the purpose of the cued retrospective report, and this method will allow for the addressing of the third goal. Given that the participants' eye movements will be recorded while taking the visual test, the researcher can explore the test takers exact eye movements, for example, examining the extent to which test takers have made eye contact to the videos and still photos during the test and also what areas have they focused on during watching the visual texts.

1.4. Significance of the study

The study intends to contribute to the field of L2 listening research by investigating the effect of the presence or absence of visuals on test takers' performance, as this aspect is still hotly debated in the literature, and still needs more profound investigation of the way test takers perceive and interact with these visual materials. Therefore, it is essential for the researcher to compare the viewing behaviour of the test takers in order to offer a detailed explanation of how different visual materials (i.e., videos and still photos) are seen by those being evaluated. This will, in effect, suggest the potential fundamental cognitive processes that are triggered when using visual materials in listening tests in addition to how this can affect students' performance in a test. More importantly, language test designers are still not encouraged to implement visual materials, especially videos, in listening tests because they do not yet know how, why and/or to what extent the use of such materials may affect test takers' performance. This is what the current study endeavours to yield through the use of eye tracking technology and the use of cued retrospective reports. In effect, shining a spotlight on this aspect could have a significant impact on our understanding of the causes of the conflicting results of research available in the field. It also attempts to create test design that may potentially be used as a start point for developing future visualised L2 listening test.

1.5. Definitions of terms

Non-verbal information:

Spoken texts convey different types of information. This information can be delivered via different channels. The most apparent is the aural channel; by which linguistic and paralinguistic (i.e., tone, pitch, stress) information is conveyed verbally (Wagner, 2007). However, another important way of transmitting information is the visual channel, by which non-verbal information is conveyed. Non-verbal information is used as an umbrella term to describe a number of features of human communication which are different from speech (Wagner, 2006b). Accordingly, non-verbal information can be defined as the components of spoken input which are not transmitted orally. These refer to features like: facial expressions, lip movement, gestures, body language, and any type of physical action.

Visual materials:

Refer to any pictorial materials that can be used to assist comprehension. It includes pictures, graphs, maps, videos and subtitles, and visual illustrations that are used to convey information visually. The term visual materials is used in this study to refer to video and still photo texts.

Video texts:

Video texts can be defined as texts that use both the audio and the visual channel to deliver a spoken input to the listeners, allowing them to hear and see the speaker's movements and reactions at the same time. The type of information conveyed by video texts can range from semantic to situational information. Video texts can be displayed live, or they can be recorded to display at a pre-chosen time. The video texts used in the current study are recorded.

Still photo texts:

Texts that convey information via the visual and aural channels and consist of either a single or multiple photographs displayed along an audio input and serve to reflect the situation in that text. Similar to video texts, still photo texts can convey semantic and situational information. The still photo texts used in the current study consist of multiple photographs that were extracted from the video texts by taking screenshots from the different recorded videos.

Audio texts:

Texts that convey spoken input to the listeners via the aural channel only. Audio texts in the current study were extracted from the video texts by separating the video from the audio input using VLC software.

Content visuals:

Visuals (can be either video or still photos.) that provide information semantically related to the spoken input.

Context visuals:

Visuals (can be either videos or still photos) that provide information about the situation or the physical setting where the spoken input takes place.

Target language use domain (TLU):

Target language use domain is simply the type of target language that language learners may deal with in their everyday, real-life situations. In the case of academic listening, the TLU is mainly the type of language that students deal with during lectures, seminars, conferences, academic workshops, etc. Linking this to test situations, Bachman and Palmer (1996) defined this concept as" a set of specific language use tasks that the test taker is likely to encounter outside of the test itself, and to which we want our inferences about language ability to generalize" (p. 44).

Chapter Two Literature review

Human history abounds in examples that highlight the importance of visual information and how they were used as the main medium to express different ideas and feelings. Not to mention that pictures found in caves and mountains from prehistory era gave us significant insight into the lives of early humans, like those drawings found in Akakus mountain in Libya (estimated 21.000 BC) which explain hunting techniques. These types of prehistory drawing have played a very important role in preserving human history. Even when the early writing system, known as Hieroglyphic, was developed, it was in the form of pictures and symbols.

The use of visual information continued to be increasingly important in the age of Islamic civilisation, where scholars like Al-Zahrawi, known as the Dean of Surgeons, who was heavily dependent in his books on using pictures and drawings to explain how human body works and how surgical tools can be used.

After that during the renaissance era, several philosophers and scholars like Rabelais, Erasmus, and Campanella called for using visual aids like pictures and maps to support learning. This idea was clearly reflected by the work of the great educator and philosopher Comenius in his “Didactica Magna” or the great didactics, which is known as the first pictorial/educational textbook. When Comenius published his Didactica Magna in the seventeenth century, he gave emphasis to the idea of “envisioning information” and described this step as extremely important for effective learning. His ideas were followed for centuries by many scholars of different educational sciences.

The idea of using visual aids has greatly progressed since those early times, and it is now increasingly used in all different aspects of life (Jordanova, 2016). In most educational fields visual materials are now being an inseparable part of the educational process (Li, 2013), and in language teaching in specific, visual aids are widely used for many different purposes, ranging from motivating students (Wagner, 2010), grabbing their attention, and helping them to retain new information (Pateşan, Balagiu, & Alibec, 2018). Visuals are also used to deliver more comprehensible input by providing illustrations, and to facilitate appropriate output by providing meaningful context that helps learners to memorize and use new information (Field, 2008; Li, 2013; Ockey, 2007; Rubin, 1994, Wagner, 2010). In our modern life, the availability of technology which offers easy use and access to visual materials in most learning contexts have increased the use of these materials in language classrooms (Li, 2013; Wagner, 2010). However, in testing contexts, visuals are still not used extensively, and that is largely attributed to the lack of empirical evidence that explains the effect of using visuals in testing conditions (Suvorov, 2013). In testing L2 listening comprehension, there is a complete absence of visual materials in almost all standardizes L2 listening tests even though most listening tests are designed to evaluate and measure the ability of test takers to perform in a target language use situation that is overflowing with visual information (Suvorov, 2009, 2013; Wagner, 2006a, 2007, 2008).

Wagner (2006) attributed the absence of visuals in L2 listening comprehension test to a number of reasons, most importantly is the lack of a widely accepted theory of listening ability. Several researchers (e.g., Bachman, 1990; Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Buck, 1997, 2001) have continuously bemoaned the absence of an accepted model of L2 listening ability that can provide a generally agreed upon definition of the construct of listening and consequently be used in creating L2 listening tests. Moreover, the lack of empirical evidence on the effects of visuals on the performance of L2 test takers, with regard to the motive that may make test takers use these visuals, in what way would they use them, and to what extent test takers utilize visualized information. Finding this out is the core purpose of this study.

This chapter will focus on the presentation of a theoretical framework for this research by reviewing the literature on areas relevant to this study, including second language listening ability, listening assessment, and the use of visual materials in listening. It will be organized into the following way: first, a general review of second language ability and its different definitions and frameworks is presented, followed by a closer look into the nature of L2 listening ability. The focus then shifts to L2 testing in general, and L2 listening testing specifically, reviewing the important aspects of listening tests and the factors that might have a direct impact on the listening process. This is followed by a review of different types of visual materials and how they have been explored in previous research, and finally concluding with the approaches used in research to investigate the impact of visual materials in listening tests.

2.1. L2 ability

Before explaining the nature of L2 listening ability, it is important to first have a bigger view of L2 ability in general. Over a long time, researchers and linguists have made several efforts to define L2 ability in a feasible and satisfactory way. Lado's (1961) attempt is recognised among the first attempts, and he defined L2 ability as a model that consists of "elements" and "skills". Each of these aspects comprises four categories, with elements consisting of pronunciation, grammatical structure, lexicon, and cultural meaning. The skills on the other hand consist of speaking, listening, reading, and writing. Reflecting the behaviourist approach to language learning, Lado (1961) stressed the importance of habit formation in the development of L2 ability, in which learners memorize sets of phrases and sentences to produce the right response to some appropriate stimuli. This approach resulted in the audio-lingual method of language teaching, which was based around the role of the teacher as the provider of correct information and the controller of the classroom. Also in 1961, Carroll made a similar attempt to capture L2 ability by defining ten competencies that were built on Lado's (1961) elements and skills model. However, when explaining L2 ability with language tests, Carroll followed a different line from Lado's discrete-point approach (which assumes that elements of language - grammatical structure, pronunciation, lexicon, and cultural meaning- can be tested separately) and called for a more integrative approach that reflect the target language use situation, where learners use the different elements of language simultaneously. Carroll also distinguished between receptive (reading and listening) and productive (writing and speaking) skills and claimed that individual language learners can vary in their levels of competence in the two.

Later, this segmentation of L2 ability into certain elements and skills was regarded as a limitation by other researchers. Spolsky in 1968 criticised Lado's and Carroll's approaches to the definition of L2 ability. He recommended that defining L2 ability should consider the fact that learning a language is fundamentally a creative process, and that what matters in that process is the extent to which learners can use their knowledge of language, not how much they know about it. The distinction between the receptive and productive skills made by Carroll was also criticised by Spolsky (1973), as he argued that "the same linguistic competence, the same knowledge of rules, underlies both kinds of performance" (p. 174). The disagreement among researchers over characterising language abilities as receptive and productive generated a bigger argument of the separability of the language skills that define the general language ability. Advocates of the general language ability factor theory claimed that different language components cannot be separable because of their interrelated nature, and the call for integrative language teaching and testing approach emerged. Oller (1979) claimed that data from language tests indicated that one factor accounted for the vast majority of the variability in language tests, and he explored the idea of a general language ability factor when he described the idea of discrete point versus integrative language tests (Oller, 1983).

However, this notion was again criticised by several researchers (Carroll, 1983; Farhady, 1983) as they claimed that this can result in a weak and non-comprehensive general factor of language ability. Vollmer and Sang (1983) suggested that language reception and production were two distinct competences, and that a two competences theory might replace the construct of a general language ability theory. Bachman and Palmer (1983) used a multitrait-multimethod matrix procedure and confirmatory factor analysis to examine different models of language ability. They argued that a partly divisible model of language ability with a general factor plus distinct traits theory was superior. Vollmer and Sang (1983) and Carroll (1983) also speculated on the possibility of a theory of language ability in which the different competences could be ordered hierarchically in some way. Because learners learn different skills and different linguistic forms at different rates, language skills “can be conceived of as being arranged in a hierarchy,” from higher order skills to more specific and explicit skills (Carroll, 1983, p. 104). Oller (1979, 1983) also referred to the theory based on hierarchically arranged linguistic components.

In 1990, Bachman used a mixed approach to reach to a workable framework to define L2 ability entitled *the communicative language ability* (CLA). This framework consisted of three major components: language competence, strategic competence, and psychophysiological mechanisms. This framework was considered as a comprehensive one as it comprised the aspect of knowledge of a language along with the aspect of the ability to use that knowledge in a communicative way. Bachman (1990) stated that this model of defining L2 ability could be distinguished from the previous ones as it comprises “the processes by which the various components interact with each other and with the context in which language use occurs” (p. 81).

Bachman's CLA model has taken different shapes since its formation in 1990. It started with the three main components: language competence, strategic competence, and psychophysiological mechanisms, and then in 1996 a modified version evolved by Bachman and Palmer. In this version, Bachman and Palmer stressed the need to view language in terms of how language learners perform language use tasks, and not to classify language use in four separate skills. Part of their argument was:

> We would not consider language skills to be part of language ability at all, but to be the contextualized realization of the ability to use language in the performance of specific language use tasks. We would therefore argue that it is not useful to think in terms of 'skills', but to think in terms of specific activities or tasks in which language is used purposefully.
> (p. 75-76)

This theoretical model is believed to present a significant framework for defining language abilities in test situations, known as the construct (Elliott, 2013; Wagner, 2006b). Before describing the components of the language ability framework, it is important to define some individual characteristics which are, together with language ability, constitute important factors that help educators understand how L2 learners use the language, and also how they might perform in language tests (Bachman and Palmer, 1996). One of these characteristics is called Topical knowledge, which refers to the use of language with reference to individuals' specific knowledge about the world (*schemata*) that is stored in the long-term memory (Macaro, Vanderplank, Graham, 2005), and specifically related to the topic in use. For instance, a geology student might find a report about volcanos easier than a student of music does. Therefore, in a standardized language test where the test is presented to different test takers from different educational backgrounds, test developers should choose general topics that do not require specific topical knowledge, as this can be a disadvantage for those who are from different backgrounds from the one in use in the test. The other characteristic is Affective schemata, which refers to the emotional connections to a specific topic. In language testing, it is important for test designers to choose topics that would not possibly bother or disturb test takers, which in turn might unfairly affect their performance (Elliott, 2013). Examples of such topic might include drugs, social problems, or racism. Topical knowledge and affective schemata are believed to interact with the components of language ability (i. e., Language knowledge and strategic competence, as explained in more details below)

Language ability was further explained in the later version of the CLA by Bachman and Palmer (1996), which consisted of a mixture of language knowledge and metacognitive strategies that make language learners capable of using the target language either in test or in the real-life situations. More specifically, the CLA framework consisted of two main components: language knowledge and strategic competence. The first component, language knowledge, is recognised as the amount of language-related information stored in memory that can be used by the language user. This component consists of two categories of knowledge: organisational and pragmatic. Each of these two categories is made up of two sub-categories of knowledge. Organisational knowledge comprises grammatical and textual knowledge. The first refers to the ability to use and understand accurate sentences, and it covers areas like knowledge of phonology, vocabulary and syntax. The latter refers to the ability to produce and understand texts consisting of two sentences or more. This involves knowledge of cohesion and rhetorical organization of oral or written texts; that is, how sentences hang together in a comprehensible discourse. Pragmatic knowledge on the other hand refers to the ability to use and understand language in real communicative situations, and relating language to its meaning, to the intention of the speaker, and to the characteristics of the context in which it was produced. Pragmatic knowledge is made up of functional and sociolinguistic knowledge. Functional knowledge refers to the ability to relate the meaning of a produced set of language to the intended meaning of its producer (intention of the speaker or writer). Sociolinguistic knowledge refers to the ability to relate the meaning of a produced set of language to the specific language use setting in which it was produced. Like using different registers according to different situation, using cultural references, or idiomatic expressions.

The second component of CLA, strategic competence, consists of a set of metacognitive strategies that work as “higher order executive processes” (Bachman and Palmer, 1996, p. 70) that control the cognitive use of language. In other words, strategic competence refers to the ability to actually utilize the knowledge of language. These metacognitive strategies are goal-setting, assessment, and planning. Goal setting refers to the process of deciding what task the language user choses from a group of tasks, in case they have the choice. With language tests, test takers do not always have the advantage of choosing a preferred task, therefore this strategy in mostly used in non-test situations.

Assessment refers to a very important strategy, by which language users relate their language and topical knowledge to the specific setting in which they are using the language. By this strategy, language users can make instant assessments to what is needed for achieving a specific task. For example, they assess the nature of a task and identify their ability to successfully achieve it, assess their language and topical knowledge to cope with the requirements of the task, and also assess the level of appropriateness of their responses to the task. The last strategy, planning, refers to the decisions made by the language users on how they will actually use their language and topical knowledge, together with their background knowledge in order to successfully accomplish the task. Bachman and Palmer (1996) provided a useful example that explains these strategies more clearly. In their example, they described a test situation where test takers are asked to choose one picture out of four pictures to describe. Test takers made different choices based on their different assessment of the requirements of the task. Because of their different assessments test takers set different goals and used different plans. For instance, the plans of some test takers involved giving a detailed description of the chosen picture, in grammatically accurate sentences. On the other hand, other test takers may set a different goal that involves describing only important information and develop a plan to use only single words in their description. It is clear from this example then how strategic competence can integrate whatever sources of information available to the language users (audio, written, visual, etc.) with their language knowledge (organisational, pragmatic) in order to utilise the language according to the specific context or situation of communication.

More specifically, with this component, CLA model presents a comprehensive and interactive picture of language ability where the learners' topical knowledge, language knowledge, and affective schemata are all interacting with their strategic competence. The strategic competence interacts with and is influenced by the dominating context or situation at the time of communication (Wagner, 2006b). In other words, strategic competence relates language ability to world knowledge and the context of the situation. Accordingly, using a language in a communicative situation would involve not only knowledge of the language, but also the ability to strategically use that knowledge in that situation. With these aspects, CLA model seems to be distinguished from the other models discussed previously, as it is the only model that practically accounted for the dynamic and communicative nature of language. The strategical use of language as highlighted by CLA framework also reflects the importance of non-verbal components of communication like gestures and facial expressions which can possibly assist the language learners' comprehension of real-life situations. Therefore, it is important that this component be considered when creating language tests. In the current study, the use of non-verbal information embodied in visual materials is recognized as an important component of assessing the test takers' listening ability, and the results of this study can expectantly inform L2 researchers on the role that visual materials play in the L2 listening process, and performance on an L2 listening tests.

Defining L2 ability is not easy as the review showed. L2 ability definitions have transformed over time, directing more attention to the importance of language in use as a communicative tool and moving away from restricting the emphasis on viewing language knowledge as purely linguistic. This as a result had its effects reflected on the way L2 listening ability is perceived, as the next section reveals.

2.2. The nature of L2 listening ability

Moving from the general L2 ability to the more specific L2 listening ability, one can clearly see that the process of defining this ability also has its own problems and complications. Despite the great importance of listening in language communication, it did not receive equal attention from language researchers like the rest of language skills, namely reading, writing, and speaking. Wagner (2008) attributed that neglect to the fact that listening was traditionally viewed as a "passive skill". What is meant by "passive" skill in this context is the ability to hear the language and memorize it in the form of specific patterns and structures related to specific situations, without necessarily being able to fully understand it and produce it spontaneously in a communicative situation (Rubin, 1994). This view stemmed from the behaviourist approach to language learning upon which the audio-lingual method of language learning was developed, which assumed that exposing learners to spoken language would provide them with sufficient input for listening comprehension (Call, 1985). Recently, more and more researchers have started to recognize the importance of this skill in language acquisition as an interactive and fundamental skill (Field, 2008; Lynch, 2011; Rost, 2013; Yeldham and Gruba, 2014) in which listeners need to use their previous world knowledge and their linguistic knowledge in order to interpret and understand the spoken message (Canning, 2002; Coniam, 2002;

Dunkel, 1991; Gruba, 1997; Ockey, 2007; Progosh, 1996; Wagner, 2006a, 2010). More specifically, viewing the role of listening has developed from perceiving it as a tool to repeat spoken input and develop better pronunciation, which dominated language teaching until the late 60's, to viewing it as an active and complex process that involves several processes that are applied in real time like discriminating between different sounds, understanding grammatical structures and recognising vocabulary, interpreting intention and stress, and linking all that with the listeners' own interpretation and what they already know about the text, as well as with the larger socio-cultural situation (Vandergrift, 2004). This understanding of listening as an interactive process has stressed the value of teaching this skill explicitly (Suvorov, 2013). With this understanding in mind, teachers can prepare students to listen by encouraging them to provide ideas related to the topic and organise them, and also triggering the suitable background knowledge which helps to make predictions. These stages of preparation for listening can largely diminish the burden of comprehension for the student.

Despite its significance as a fundamental aspect of language acquisition, L2 listening comprehension is still largely under researched (Bloomfield, et al. 2010; Walker, 2014), and that can be attributed to various reasons, like the inherent difficulty in providing a complete and thorough definition of L2 listening (Wagner, 2013), and the dynamic and ephemeral nature of listening which makes it one of the most difficult skills to measure (Flowerdew and Miller, 2005; Ockey and French, 2016)

The complexity of listening can be explained by the fact that listeners during their reception of spoken messages, need to integrate information from multiple sources: phonological, lexical, prosodic, syntactic, semantic, and pragmatic. All these sources need to be processed automatically and in real time (Buck, 2001; Celce-Murcia, 2002). Rost (2014) has also stressed this fact by depicting listening as a combination of neurological, linguistic, semantic, and pragmatic processing. These factors are a reflection of the complicated nature of listening, which explains the difficulty and the challenging nature of assessing it. This could also explain the dearth of research that listening has received compared to other skills, especially, research on listening assessment. Therefore, before any attempt to discuss and understand the construct of L2 listening, it is important to first elaborate more on the processes and variables related to listening in order to be able to define and understand its construct. The following section provides more details into the process of L2 listening ability.

2.2.1. Models and processes of L2 listening comprehension:

Beside the numerous definitions of listening, several researchers (Bejar, Douglas, Jamieson, Nissan & Turner, 2000; Buck, 1991; Weir, 1990) have also proposed models for listening comprehension in an attempt to better understand the cognitive processes that underlie listening.

Listening has been traditionally described as a two-stage process. Many researchers (e.g., Buck, 2001; Field, 2008; Lund, 1991; Lynch, 2009; Rost, 2013; Secules, Herron, and Tomasello, 1992; Vandergrift and Goh, 2012; Weir, 1993) have studied the idea of dividing the cognitive processes of listening into two levels (generally speaking, a decoding stage and an interpreting stage), in spite of using different terminologies for the two processes or stages. These processes as a whole have been defined by different researchers. One of the simple descriptions of these processes has been presented by Buck (2001) who depicted it basically as a first stage that extracts important linguistic information and a second stage that uses the extracted information for communicative processes. However, Buck's description of the listening process seems to be oversimplified and shows listening as a straightforward, linear process.

Buck has referred to a number of researchers (e.g., Carroll, 1972; Clark and Clark, 1977; Rivers, 1966) who have also explained that listening comprehension is composed of two-stage processes as well. In (2001) Buck stated:

> These scholars seem to have arrived at similar conceptualisations of listening comprehension, and the fact that they use different terminology suggests that they have arrived at this understanding more or less independently. This adds considerable credibility to the two-stage view of listening. (p. 52).

What these researchers have suggested, is that the variety of listening sub-skills is attributed to these two different stages. That is, they referred to those sub-skills that feature decoding process as typical of the first stage, and sub-skills that feature interpreting process or constructing meaning as typical of the second stage (Wagner, 2006a). Among the different descriptions of this two-stage process, most researchers refer to top-down and bottom-up processing. Some researchers have provided a simple analysis of the way these two processes work. For instance, Kelly (1991) has described bottom-up processing as the process in which listeners receive the spoken input as sound (i.e., by combining phonemes, syllables, words, and clauses) and then start to interpret its meaning. The top-down process, on the other hand, indicates applying cognitive processes that depend on the prior or world knowledge to the sound input in order to give it a comprehensible meaning. Kelly (1991) has explained that the sound input serves to confirm whatever expectations the mind made. This implies that perception occurs when enough information from both sources is presented and processed, and in this case the two processes cannot be linear.

Similarly, Peterson (2001) suggests that lower-level processes (bottom-up) serve to extract meaning from sounds, intonations, words, which once identified can fit into the larger units, like phrases and sentences to match related information stored in long term memory, which is normally known as schemata. Prior knowledge or schemata refers to the complex mental knowledge of the world that is acquired by the listener at any point in time (Macaro et al., 2005), hence, it refers to the type of information stored in the long-term memory about various topics, cultures, ideas, and other sorts of information.

Rost (2006) has also provided a model with rather simple explanation of the two processes as the bottom-up processes referring to speech perception and word recognition at the linguistic level that supply the listener with data necessary for comprehending the spoken input. If, according to Rost, the listener cannot recognize a sufficient amount of these data for any reason, she or he will automatically depend more on top-down processes, such as, semantic expectations, inferencing and generalizations in order to compensate for the lack of linguistic knowledge and better understand the text. Top-down processes in this sense are related mainly to the schemata or the prior knowledge that the listener has. Rost’s idea of how listeners sometimes rely more on one process (top-down) in order to compensate for vocabulary insufficiencies complies with what has been argued by some other researchers (Bloomfield, Wayland, Rhoades, Blodgett, Linck, and Ross, 2010; Goh,
2000; Tsui and Fullilove, 1998) as will be explained in more details in the *characteristics of the listener* section below.

However, this theoretical distinction or separation between these two cognitive processes seems to be simplified and to some extent unrealistic in this way, since it implies that the two processes are dissociated from each other, and that they occur in a successive manner, without any kind of interaction between them. This led some researchers to propose new models in an attempt to improve the two-stage process model and to understand the mechanism of listening processing. For instance, Flowerdew and Miller (2010) in their suggested model argued that listening comprehension is composed of three main cognitive models instead of only two. These are bottom-up model, top-down model (which are the same as explained previously), in addition to the interactive model. This later model consists of a combination of the two previous models, and views listening as a "simultaneous interpretations of auditory input at different levels" (p. 170). This includes a combination of phonological, syntactic, semantic, and pragmatic information.

In fact, it seems that there is a tendency among researchers to refine the theoretical relation between the two stage processes instead of adding a new one. Field (2013), for instance, has explained the more complex nature of the two cognitive processes when he explained that the bottom-up process is not a fixed process that goes through all the early mentioned stages of combining phonemes into syllables, words, clauses, etc. instead, he argues that the bottom-up process does not always involve all these levels. For example, in some instances, listeners may process words from phonetic features directly without going through the syllable level, which implies that the listener can process words even before the speaker finishes saying them. Accordingly, hypotheses start to form as the word is being pronounced. These hypotheses are activated to various degrees until one of them matches the sound signal and the rest will be discarded. Sohoglu, Peelle, Carlyon, & Davis (2012) have also attempted to investigate the nature of interaction between these processes, and they stated that "A remarkable feature of the brain is its ability to integrate these two sources of information seamlessly in a dynamic and rapidly changing environment. However, the mechanisms by which this integration takes place are still unclear" (p. 8443).

Similarly, Wagner (2007) argues that both types of processing, bottom-up and top down occur concurrently and interactively instead of serially. The two types influence and affect each other, although they might not be utilized constantly equally by the listener. Likewise, Vandergrift (2015) asserts the interactive and interpretive nature of listening processes where the two types of processes are employed “simultaneously and in a parallel fashion” (p. 300) to comprehend the message. This means that while a person is listening to a spoken input, one type of processing might be controlling at certain times, and the second type is at control at other times, depending on the purpose of listening and also on the type of tasks.

Background knowledge, familiarity with vocabulary, topics, and contexts, are all factors that affect the domination of one type of processing over the other. For instance, Vandergrift (2004) illustrates that when the listener is listening for the main idea or the gist, top-down processes will be involved more. While listening to specific details mainly involves bottom-up processes. Similarly, Macaro et al., (2005) have implied that top-down processes are usually activated with specific topic texts to which the listener has a high degree of prior knowledge, while bottom-up processes are typically associated with low prior knowledge topics. Kelly (1991), in an attempt to explain this interaction, has claimed that listeners' predictability and familiarity of words and texts will lead them to rely less on explicit bottom-up processing. While in other cases where the listeners' meaning expectations are low, they may find themselves obliged to highly use the sensory level of bottom-up processing. Kelly (1991) has also stated that because it is difficult for beginner learners to predict the meaning of words and texts, they usually have low meaning expectations of the forthcoming spoken input; therefor they might be forced to depend largely on bottom-up processing.

However, other researchers (Goh, 2000; Rost, 2006; Tsui and Fullilove, 1998) have different views. They argue that because low level listeners do not know the meaning of many spoken words and their vocabulary knowledge is relatively sparse, they will compensate for this lack by relying more on top-down processes to infer the meaning of the message.

In terms of the use of visual materials with listening, it seems that providing a visual aspect would most likely assist the listeners’ top-down processing of the audio input. More precisely, according to Yeldham and Gruba (2014), that among the traits of the top-down processes is the capacity to guide, contextualize and enrich the linguistic input in order to facilitate its interpretation. Accordingly, it can be argued that providing the visual input would support the top-down processes and act as an important facilitator to the process of interpreting the linguistic input, as it enriches that input and provides the listener with a clear idea of the context in which the speech took place, which may help them to achieve better comprehension.

Another important explanation of the process of listening was given by Wolvin and Coakley (1996), who defined listening as “The process of receiving, attending to, and assigning meaning to aural and visual stimuli” (p. 69). This definition seems to tackle the process of listening step by step and breaking it down into three main components. Starting with “receiving”, this part of the definition refers generally to the journey of sound waves (and visual stimuli) to the brain where it will be translated. Next, “attending to” refers to intentionally focusing the perception on selected stimuli by filtering out other noises and sources. This is an important component in listening, because humans have limited cognitive processing capacity, the listener can only handle a limited number of sources of information at the same time. For instance, when the listener receives more than one stimulus at a time, they will choose to focus on one of them to attend and listen to and neglect the others, otherwise the listener will end up with unclear glimpses from the available stimuli. The attending behaviour therefore refers to the fact that the act of attending to the message is controlled by the listener, at least to some extent (Wagner, 2006a). However, it can be argued in this case that when the different stimuli available do complement each other, the listener would possibly find them helpful to comprehension and might choose to attend to them all simultaneously. The last component in the definition, assigning meaning to stimuli, involves multiple processes in itself. It refers to understanding, remembering, and responding to the message. As Wolvin and Coakley (1996) upgraded their definition of listening from “assigning meaning to aural stimuli” (1988), to “assigning meaning to aural and visuals stimuli”, it came as they realised that when those two channels (aural and visual) are combined, they will have more impact on assigning the meaning, remembering, and responding to a stimulus than attending to only one of them (Schnapp, 1991). Hence, the act of listening is associated with comprehending the meaning of the verbal or auditory information presented, and consequently responding to these perceived stimuli.

Based on these processes, Wolvin and Coakley (1996) created a model composed of two-part process where the top part represents the listener in a communication interaction, and the bottom part represents the same person once they switch to the responding stage and becoming the speaker. According to Wolvin and Coakley (1996), the process of listening ends once the person attempts to respond, as they then assume the role of the speaker.

As there are other earlier models suggested by different researchers like Bostrom (1990) and Wolff et. al, (1983), almost all these models have similar structures and processes to the ones created by Wolvin and Coakley (1996). Nonetheless, most of the models discussed in this chapter share two major limitations:

The primary limitation of cognitive listening models is that none of them have been proven through empirical validation. Despite attempts to validate the Wolvin and Coakley (1996) model, such efforts have been unsuccessful, as evidenced by Janusik and Wolvin's (2001) research.

A second limitation is that none of the models have been able to effectively differentiate between listening and cognitive processing. Cognitive processes were generally viewed as one component of Listening, rather that the entire process itself.

The above review of listening models and processes indicate that there are still limitations and differences among researchers, and there is not yet a unified model for listening comprehension. As a result, no generally agreed upon theory or model of listening exists (Buck, 2001; Suvorov, 2013). This diversity in researchers’ views of listening can be attributed to the different definitions and also to the different purposes of listening (Olson, 2003) which are used as foundation to the various classifications of listening types.

2.2.2. Defining L2 listening ability

As described above, listening can be roughly defined as the act of processing auditory and non-verbal information of spoken text in order to construct meaning from that text (Wagner, 2006b). However, despite how logical this definition seems, an adequate and consistent definition is still proving to be difficult (Bejar et al., 2000; Bloomfield et al., 2021; Buck, 1994, 2001; Lynch, 1998; Vandergrift, 2015; Wagner, 2010), and that resulted in a lack of commonly accepted definition of L2 Listening.

Wagner (2006b) explains that providing a definitive and comprehensive definition is largely elusive because there are many different psychological and cognitive variables and processes involved in L2 listening ability which vary according to the different listening situations. For instance, Richards (1983) explained how the processes of L2 listening change according to the type of the listening material (academic listening, social interaction, listening for pleasure, listening for information, etc). Likewise, Buck (2001) claimed that the manner of processing aural input by L2 listeners varies according to the different context of the situation. In addition, the processes used in L2 listening vary according to the listener's ability level. This means that L2 listeners with lower ability level process audio texts in a different approach than L2 listeners who have higher ability level (Buck, 1994).

Nevertheless, despite the difficulties, many researchers stress the importance of defining listening adequately and uniformly before any attempt to study it (Witkin and Trochim, 1997), and they see that this is the only way L2 listening research can proceed. As this still proves to be difficult to attain, some other researchers think that since listening is multidimensional in nature, it would be better to have multiple definitions according to its different purposes and uses (Bodie, Janusik, and Välikoski, 2008).

Accordingly, the focus has shifted from focusing solely on linguistic knowledge to a greater emphasis on language as a tool for communication and use in context. As advised by Wagner (2010), a comprehensive definition of L2 ability should highlight the importance of language use, context, and communication, and take into account the important role of pragmatic ability and cross-cultural differences on appropriate language use. It should also recognize the interrelationship of different linguistic components of the language in a hierarchical order and acknowledge the dominance of certain language forms and functions. This in fact, has opened the door to including other elements to the listening ability (Ockey and Wagner, 2018), and led to the growing agreement among researchers about the importance of the non-verbal, visual element in listening. For instance, since 1995, Rubin has defined listening as "an active process in which listeners select and interpret information which comes from auditory and visual cues in order to define what is going on and what the speakers are trying to express" (p.151). Moreover, Wolvin and Coakley (1996) have described listening as "a process of receiving, attending to, and assigning meaning to aural and visual stimuli" (p. 69). This definition of Wolvin and Coakley is in fact a refined version of their earlier definition in 1988, which did not include the *visual stimuli* component as mentioned earlier. This modified definition of listening refers to the continuous evolution in the understanding of the process of listening.

In testing situations, it is essential to provide a clear definition of the type of listening that is going to be tested which states all the abilities that the test designer wishes to test, which is known as the construct. Before attempting to do so, it important to know first how different researchers classified listening according to different purposes. These different classifications resulted consequently in the formation of various L2 listening constructs by test designers. The next section reviews these different classifications and taxonomies proposed by different researchers, followed by an overview of how these differences in classifying listening types affected the formation of a construct for L2 listening comprehension.

2.2.3. Taxonomies of listening

The difficulty of providing a comprehensive definition of L2 listening resulted in several classifications and taxonomies to describe listening comprehension skills, each of which is based on different purposes for listening. The result is varied classifications, some of which are simple and straightforward, others are rather complicated. For example, Richard (1983) perceived listening as a number of micro-skills, then he grouped these into two main categories. First is Conversational listening which includes 33 micro-skills like; the ability to process speech at different speeds and the ability to retain chunks of different length of spoken sentences. Second is Academic listening, and that includes 18 micro-skills like; the ability to identify the main topic of a lecture and to infer meaning of words from the context. In 2007, Brown proposed a simplified version of Richards' taxonomy, in which named 7 skills involved in Academic listening and classified them as macro-skills, as they are used in discourse level, while listed 10 micro-skills for conversational listening and those are the skills which are used in the sentence level. Buck and Tatsuoka (1998) somehow used a similar taxonomy in terms of classifying skills into two main groups. They grouped 15 prime attributes like; the ability to understand and correctly use stress patterns and the ability to process fast spoken speech. The second group of skills consisted of 14 interaction attributes which include skills like; the ability to make inferences from the context of speech. Another taxonomy was presented by Field (2008) which to some extent is similar to the previous classifications, but this one is perceived from the listeners side, which mean that it is based on the listener's goal. Accordingly, he presented two main categories for listening: local and global. The first refers to actions that usually extract details, like focused and unfocused scanning (locating specific information needed by the listener), message listening, search listening, etc. the second, on the other hand, refers to extracting the gist, like listening for plot (e.g., films or TV dramas), skimming, listening to check critical facts, etc.

There are other taxonomies which differed from the previous ones by classifying listening into more than two groups. For instance, Bejar, et al., (2000) have classified listening into four types: (1) Listening for basic information. (2) Listening for specific information. (3) Listening to integrate information. (4) Listening to learn. Also, Wolvin and Coakley (1996) have identified five types of listening: (1) Discriminative listening: by which the listener discriminates between the verbal and non-verbal message (e.g., a friend says he is happy, but his face expressions are sad). (2) Comprehensive listening: which is listening to understand the speech (e. g., understanding instructions for accomplishing a task). (3) Appreciative listening: listening to topics that are highly entertaining, this includes listening to music, television, etc. (4) Empathetic listening: is the ability to listen to other peoples' feeling and being able to show sympathy with them. (5) Critical listening: the ability to evaluate and judge the message.

Hence, it is clear that listening classifications differ greatly according to different researchers, and no two classifications are the same. As a matter of fact, not only are the classifications of listening different, but some researchers have also used different terms in classifying listening. For instance, the term "skill" has been used by Brindley (1998) as in the skill of understanding main ideas, the skill of listening for specific information, and the skill of inferencing the speakers' meaning. It seems that by using the term skill, Brindley intends to refer to these types of listening as an ability that can be taught explicitly.

Here, some researchers suggest that academic listening cannot be bounded to a fixed type of listening, since many other types of listening (as defined previously by a number of researchers like: Bejar et al., 2000; Field, 2008; King and Behnke, 1989) can also occur in the academic context. Therefore, some researchers have argued that academic listening should not be considered simply as a type of listening; instead, it can be regarded as involving a specific register of language (Suvorov, 2013). Here, *register* can be basically defined as the variation of language according to the function it serves. In other words, the use of language differs according to the *field*, *toner*, and *mode* (Halliday and Hasan, 1989). *Field* refers to the theme of communication (e. g., business meetings). *Toner* refers to the nature of relationship between participants (e. g., formal, informal/ requires respect). *Mode* refers to the means of communication (e. g., spoken, written, or non-verbal). All these variables control the choice of register and the variation of lexicons and grammar applied (Watts, 2001).

It is obvious therefore that the process of classifying types of listening is not a straightforward and unified one. These variations were also reflected in the field of listening assessment and affected the development of L2 listening constructs. The next section explores listening assessment in more details, highlighting the importance of defining the
L2 listening construct and explains the challenging nature and complexity of this process.

2.3. General background on language testing

Language testing in general shares a common foundation: its usefulness (Bachman and Palmer, 1996). As the primary purpose of any test is to measure abilities, by making inferences about those abilities from the scores of test takers, it is then essential that the test should include certain qualities that can make these inferences valid. Six qualities established by Bachman and Palmer (1996) are: reliability, construct validity, authenticity, instructiveness, impact, and practicality. Generally, these concepts can be defined in the following way*: Reliability* is known as the extent to which the measurement of a certain test is consistent across different situations and with different examiners. *Construct validity* is the extent to which the interpretations of the scores of a given test can indicate the abilities specified by the construct (construct is defined in more details in section 2.3.2). *Authenticity* is the extent to which the characteristics of a given test correspond to the features of the target language use domain. *Interactiveness* is the extent to which test takers' language ability, topical knowledge and affective schemata interact to accomplish the test tasks. *Impact* refers to the influence of a given test on the individual, the educational system, and the whole society. *Practicality* is defined as the procedure of making use of resources in order to develop and use the test in a practical way. Of importance to the current study are Construct validity and Authenticity. That is not to say that Reliability, Interactiveness, Impact and Practicality are not important factors, in fact reliability is critical to language tests and considered as an essential quality. However, the scope of the current study does not cover issue of controlling inconsistences of test scores, design of test tasks that influence interaction with it, or the ultimate effect of the test on individuals and the practicality of it. The following section focus on L2 listening tests and introduces the two test qualities related to the current study.

2.3.1 Qualities of L2 listening assessment

Testing listening comprehension has gone through different stages since the beginning of the practice of L2 language assessment in the beginning of the last century. The early attempts of assessing listening ability did not contain a designated part for listening, instead language tests most often included a dictation section, a phonetics paper, in addition to a conversation section. The outcome of all these components point to some evidence of the ability to listen (Taylor, 2013). Over time, continuous revisions of the quality of L2 language tests led to dropping the phonetics paper while keeping the dictation section. During the 1960s, dictation was also replaced, and listening was delivered by an examiner who would read aloud a passage to the test takers. During the 1980s and 1990s, with the rapid advances of technology, cassette recorders became available, and that affected the delivery of listening test, as the human speakers were replaced by recorded materials. Cassette recorders were also replaced latter by compact discs (CD), which provided better sound quality. As technology continued to advance, computer-based tests started to appear alongside the traditional paper-based tests. The multimedia capacity of computer-based tests opened the doors to including visual materials (Elliot and Wilson, 2013), and that as a result encouraged test designers to explore and create new ways of delivering different test, especially listening. This in turn required reviewing and enlarging the way L2 listening construct was perceived.

2.3.2. The construct validity of L2 listening.

In the field of listening assessment, test developers constantly face the problem of providing a construct for the listening ability to be used as a base for developing and validating the test.

As a starting point, what is meant by a construct needs to be clarified. Buck (2001) has provided a straightforward definition of a construct as "the thing we are trying to measure" (p.1), he also added that a test will be valid only when it measures the right construct. In other words, if the construct was valid, the inferences made by test developers about the test takers' scores can be valid too. Therefore, test developers should always provide a clear construct definition of the ability they are trying to measure. According to Buck (2001), this includes two stages: first, defining the construct theoretically, and second, operationalizing it through texts and tasks selection. Sometimes in the stage of operationalisation, test-developers might add an unrelated variable or omit a fundamental one. This is what was explained by Messick (1996) as *construct-underrepresentation*, for omitting important variables included in the theoretical description of the construct, or *construct-irrelevant variance*, for assessing abilities which are not included in the theoretical base of the construct.

It has also been argued that an appropriate construct-definition that can form the basis for a reliable and valid L2-listening test is governed, in part, by who the test takers are, their age, what their purpose for listening is, and how they listen. Defining the L2listening construct is a large and complicated process that involves numerous factors.
According to Field (2012), if test scores are used to predict language ability (as TOEFL and IELTS test scores often are), then the test developers should design a test that has the test takers use the language in the way they will probably need to use it in their future situations.

Several researchers suggested different approaches to define the construct of listening, as well as presenting several types of knowledge and skills which are required for L2 listening performance that can be included in the construct. In fact, many of those researchers have a congruent view about the importance of visual cues in listening comprehension (Baltova, 1994; Rost, 2013; Sueyoshi, and Hardison, 2005; Suvorov, 2013; Wagner, 2006b). However, there is no total agreement yet about the significance of including the ability of understanding visual information in the construct definition of listening. For some language testing researchers (Buck, 2001), language is viewed as consisting of four separate skills (listening, reading, writing, and speaking) and each skill should be measured and assessed separately from the others. This view implies that listening tests should employ the audio input only.

Nevertheless, as the field of L2 listening assessment has greatly developed over time, its level of complexity has also grown as a result (Vandergrift, 2007). Recently, there has been a growing awareness among language testing researchers about the need to revise and enlarge the scope of the L2 academic listening construct by comprising the ability to understand visual information in it (Ockey, 2007; Wagner, 2006b, 2007). Aryadoust (2022) stresses that the aspect of visuals should be emphasised in the construct of L2 listening. He explains how language for academic purposes is often very closely connected with visuals and giving examples of how using techniques like PowerPoint slides, short video clips, still pictures, or even simple things like the lecturers' body language, are inseparable components of academic listening. Aryadoust (2022) states that "decontextualizing L2 listening by excluding visuals would divest it from its authentic features" (p,5). Similarly, Wagner (2006b) has considered that ignoring this important type of knowledge in the construct can lead to construct underrepresentation. Okcey (2007) has also recommended revising the construct of listening when using video texts, to include the *ability to use visual information* in order to support listeners' comprehension. These abilities are explained by Ockey as understanding hand gestures, lip reading, and facial expressions' interpretations. They also refer to the ability to recognize information provided by context/content input, like the place where the lecture took place, the illustrations provided by the lecturer (e. g., diagrams, pictures, charts, maps, objects related to topic, etc.) and the whole surrounding environment of the lecture.

Consequently, including the ability to understand visual information in listening assessment by developing visualized texts seems to be increasingly dominating. For instance, large-scale language tests such as TOEFL IBT are increasingly incorporating integrated skill assessments, indicating a shift away from the traditional belief that language is comprised of separate skills. This trend recognizes the importance of a multidimensional approach that includes various abilities specific to the language task at hand, as outlined by Bachman and Palmer in 1996. This approach goes beyond verbal comprehension to encompass the interpretation of non-verbal cues such as facial expressions and gestures, as highlighted by Ockey in 2009. Moreover, vision normally works as a part of a larger system, in combination with hearing, smelling, tasting, and tactility (Zielinski, 2006). Hence, visual information is closely related to verbal information, and they enrich each other in most communicative situations. In addition, language is being increasingly viewed as an interactional system of verbal and visual modes (Royce, 2007), as the growing interest in multimedia learning and studying multimodal communication competence demonstrates (Mayer, 2005). This implies the significance of including visuals in contemporary use of language and in constructs used in language assessment.

For the current study, the construct for L2 academic listening is developed as: the ability to process and comprehend information provided by acoustic and visual input from academic lectures in university settings where English is the main language of communication. The use of visual materials in L2 academic listening test is regarded as an important reflection of the target language use domain, where in almost all academic lectures students can see the lecturers and any supporting materials they are using, like smart boards or Power Point slides. Therefore, understanding the visual aspect is included as one component in the L2 listening construct of this study alongside the verbal aspect. This combination of different sources of knowledge is supported by the theoretical model of communicative language ability as proposed by Bachman and Palmer (1996) as described previously, which best represents the interaction between language knowledge and the ability to use that knowledge in test situations as well as in real-life situations. More specifically, it is believed that the ability to construct meaning from the audio and visual resources involves the ability to use or invoke strategic competence to integrate the incoming aural and visual information with the linguistic and non-linguistic knowledge of the listener, and the context of the speaking event.

2.3.3. Authenticity of L2 listening texts

Authenticity is closely related to construct validity, as it refers to the degree of correspondence between the characteristics of the test and the characteristics of the target language use domain (Bachman and Palmer, 1996). Authenticity is used here to refer to the listening texts rather than to test tasks. However, in many studies, the concept of authentic texts in listening comprehension referrers to the use of materials from real sources that are not designed for test purposes. For instance, a general definition of authenticity of the text is provided by Gilmore (2007) as “A piece of real language that is created by a real speaker for a real audience in order to convey a message of some sort” (p. 97). This sort of definition seems to be very general and absolute, and that led some language teachers and test designers to select pieces of spoken texts either from the radio or TV and including them in their L2 listening test as authentic materials (Safarali and Hamidi, 2012). In language testing situations, authenticity is considered as a critical parameter that is inseparable from all consideration of validity (Elliot and Wilson, 2013). Field (2013) asserts that authenticity should not be associated with the production and use of authentic texts “for its own sake” (p. 110), but rather with the extent to which these materials reflect the level of difficulty that would be faced by test takers in their target language domain in the rea-life. Similarly, Taylor (2013) stated that full authenticity cannot normally be achieved in test situations. Instead, test designers should attempt to create tasks and situations that would correspond to the level of language and tasks which most likely be faced by the test takers of a specific domain.

One of the main reasons for considering authenticity in language tests is its potential influence on test takers’ perceptions about a given test, which in turn affects their performance (Wagner, 2007). More precisely, the degree a language test is relevant to the features of the target language use domain, by choosing suitable topical content and appropriate task types can probably affect test takers in a positive way and promote the way of answering the test tasks, and eventually improve their performance (Bachman and Palmer, 1996). Therefore, the first step in the attempt to achieve an acceptable level of authenticity in any language test is to define the unique characteristics of the TLU, to which the test text should correspond.

In terms of the current study, and as stated in the previous section, the TLU is academic lecture listening where visual materials constitute a vital component which in the majority of cases is inseparable from the spoken input. Hence, the visual materials used in this study are designed to mirror natural academic lectures in order to achieve a suitable level of authenticity that can potentially promote test takers perceptions and performance. In support of authenticity by using visual materials in the texts of L2 listening tests, Field (2012) argues that it is advisable to strive for authenticity by means of including visual materials if the test reflects a target language domain where the listener is either an active or passive participant with the visual situation.

In sum, the ability to understand visual information can be argued to be an integral part of language comprehension, yet it is not fully employed in the field of second language assessment in general, and language for academic purposes in specific. Since academic listening is employed in the current study; a closer look is needed to explain the nature of this type of listening.

2.3.4. Academic listening

Academic listening was traditionally viewed in the literature as an entirely different from conversational listening in terms of the skills needed in each one (Flowerdew and Miller, 1997; Richard, 1983). In most academic settings, students learn through a number of methods by which they gain knowledge. Commonly, there are three main methods: lectures and presentations, interactive activities, and project work (Hillier, 2002). This means that academic listening can be either message oriented, which is also known as one-way listening or transactional listening, or it can be people oriented, which is in this case known as two-way listening or interactional listening where the listener interacts with at least one other person (Gu, 2018; Lynch, 2011; Suvorov, 2013).

Higher Education Policy Institute (HEPI) estimates that students spend on average about 13.1 hours a week with lectures only, which is roughly half of the average total of time-tabled learning (Shepherd, 2012).

The focus in this study will be on transactional listening as in academic lectures, which is described in the literature as a message-oriented use of language and functions mainly to communicate, describe and provide information, give directions or instructions, explaining, requesting and confirming understanding (Morley, 2001). According to Vandergrift (2004) transactional listening (e. g., lectures) requires listeners to keep paying attention to a larger degree than interactional listening, because they usually do not have the chance to communicate with the speaker and ask for clarifications or repetition.

However, this seems to apply for recorded materials where the speaker is not physically present with the listener, since in most real lectures the listener at some point can ask the lecturer to clarify ambiguous points, although it is not common to ask for interpretation of specific word meaning and that is probably what Vandergrift (2004) has pointed to.

In academic listening, L2 learners will need to listen to a large amount of information in a relatively long period of time, about 50 minutes (Jeon, 2007). This requires learners to train and develop their comprehension strategies, like the ability to focus on key points in the lecture and ignore other minor points, and improving their inferencing ability, whether through auditory or visual input (Suvorov, 2009; Wagner, 2006a). They also need to master a number of micro-skills like real time processing of the spoken input, and at the same time they might even need to take notes (King, 1994). In addition, L2 learners in academic settings need to apply their background knowledge (activating their schemata) for topics in the specific field of their study. According to Y'ian (1998) and Tsui and Fullilove (1998), the non-linguistic knowledge needs to be activated by the linguistic knowledge. In other words, in order to activate the background knowledge, listeners need as a starting point to process the linguistic input (bottom-up processing), in which they need a rapid and accurate access to word meanings (Jeon, 2007; Tsui and Fullilove, 1998). In fact, this could be a major challenge for L2 learners in particular, since it is totally different from L1 academic settings where language processing is not an issue and happens automatically (Yi'an, 1998). To summarize, the key features that distinguish academic listening from other types of listening (e. g., conversational) can be specified as follows:

- Normally requires specific subject matter knowledge.
- Does not require turn-taking as in conversational listening.
- Academic speech is normally informative and direct.
- Requires concentrating on long stretches of speech.
- Requires the ability to take notes.
- Requires ability to integrate incoming verbal information with other media (e. g. textbook, handouts, or visual materials like PowerPoint slides)

These features which may be challenging to L2 listeners have led many educators and researchers (Ockey, 2007; Sueyoshi and Hardison, 2005) to recommend the use of some strategies to increase the quality of the academic input and support comprehension, like using technology in order to help students to better understand and digest information. One of these improvements is the increasing use of visual aids to accompany the audio message. Nowadays the use of visual support is becoming an everyday procedure in the practice of academic teaching because educators (Ockey, 2007; Suvorov, 2013; Wagner, 2006a) believe that learners will benefit more by integrating information that come from the audio as well as the visual modes, like the use of Power Point, diagrams, pictures, etc. In addition, Baddeley (1992) believes that the human brain has independent processing streams for visual and verbal information. According to his research results, Baddeley argues that dualchannel processing of information is better than single channel, and that learning can be enhanced when instruction involves both visual and verbal information (Mayer, 2005).

It is then evident how important the use of visuals for academic listening is. Yet, when it comes to the assessment of academic listening, visuals are still challenging to L2 test designers and experts. In fact, part of this challenge is that test developers must consider a list of variables that can affect performance in listening tests. Some of these factors are discussed below.

2.3.5. Factors affecting performance in L2 listening comprehension tests.

L2 Listening comprehension does not only involve cognitive processing of the incoming spoken input; it can also be affected by several other factors like the listeners' motivation, memory capacity, and the quality of the spoken text itself. As yet, little is known about these factors and their impact on L2 listening ability (Vandergrift, 2015), thus scholars have addressed them slightly differently, but the main factors remain to a great extent the same. Here, a list of some of the factors affecting L2 listening comprehension is presented, as seen to be related to the current study:

- Characteristics of the listener (i.e., L2 proficiency, metacognitive strategies, anxiety, and working memory)

- Characteristics of the text or input (authenticity, text complexity, length, and organization)
- Characteristics of the speaker
- Characteristics of the test condition (time limits and associated response).

Characteristics of the listener:

There are many individual differences among listeners that may have an impact on listening comprehension to different degrees. Some of these factors are discussed here as they are considered to be related to how the listener perceives the spoken message and deal with task difficulty. These factors include working memory, L2 proficiency, use of metacognitive strategies, and anxiety.

Working memory

Working memory can be defined as a cognitive system that is responsible for processing, storing, and retrieving information in memory (Gathercole and Baddeley, 2014). It is believed that working memory consists of two main components: storage and attentional control (central executive). Recent research related to working memory emphasizes the importance of the central executive as a main determiner of individual differences in working memory, and it is also responsible for managing attentional resources (Engle, 2002; Gathercole and Baddeley, 2014). This points a finger to the critical role of the central executive component in affecting the level of L2 comprehension.

Most existing research on working memory and L2 comprehension is done on reading. However, as the inner cognitive processes of reading and listening are similar in general, Bloomfield et al., (2010) argues that many of the results on the role of working memory in comprehension can be generalized from reading to listening. The most important difference that should be considered is the ephemeral nature of listening compared to reading. One cannot go back to previously mentioned points in the spoken message, which can affect working memory by imposing extra load on it in the case of listening.

The little available research on the effect of working memory on L2 listening comprehension suggests that processing L2 input places an extra load on working memory

(Vandergrift, 2015) which may, in some cases, lead to incorrect judgments about spoken L2 input, for instance; challenging grammatical structure of the input can confuse the listener even if the vocabulary used in the text are fairly simple and intelligible (McDonald, 2006). In terms of the effect of working memory on L2 listening supported with visual materials, research seems to be in its very early stages and much more attention is still required. Vandergrift (2004) however claimed that the inclusion of visuals in L2 listening might result in limiting the amount of available working memory by consuming the attentional

resources, and that can distract the listener from paying attention to the required information. This claim by Vanergrift (2004) was not based on empirical evidence, as he was commenting on the results of Ginther's (2002) study on the effects on including visuals on the performance of test takers in TOEFL listening.

L2 proficiency

Proficiency is an important factor in L2 listening comprehension, although the lack of an adequate, standard definition that determines proficiency levels (low, intermediate, and high proficiency levels) can affect the accuracy of interpreting research results. Proficiency level is usually investigated in the literature through categorizing it according to three main types of knowledge: vocabulary knowledge, phonological knowledge, and background world knowledge. These types are briefly discussed below with reference to listening.

- L2 Vocabulary size: refers to the amount of vocabulary or the number of words that listeners know. Once again almost all measures of vocabulary size and comprehension have been done on reading (Lynch, 2011). However, it is estimated that listeners need to know about 5000 words (the most frequent) in order to understand the main points and general idea of a text (Bloomfield et al., 2002). However, the ability to manage difficult and unknown L2 vocabulary among listeners differ greatly (Vandergrift, 2015), and these differences might be related to the listeners' metacognitive ability, as will be explored in more details in Metacognitive strategies section below.

- Phonological knowledge: most research on phonological knowledge has focused on the relationship between listeners of different proficiency levels and the type of processes they rely on most, bottom-up or top-down. The major results of these studies are mostly consistent and suggest that less experienced listeners usually depend more on top-down processes in order to compensate for the unknown vocabulary or filling in missing details. This is because their awareness of the target language phonology and knowledge of grammatical structure are not sufficient yet, and that can lead easily to missing the meaning of a large number of words in the spoken message (Goh, 2000; Tsui and Fullilove, 1998). Therefore, those low-level learners tend to compensate for their phonological and grammatical deficits by relying on their background knowledge to understand the spoken input (Bloomfield et al., 2010).

- Background knowledge: or Schema as it is known in the literature, can be defined as a cognitive framework that works as a device to interpret and organize incoming information, and it can also provide shortcuts to previous existing information and idea (Long, 1989). Some researchers do not include background knowledge as a component of L2 proficiency and classify it as a separate category of listener characteristics. However, following Bloomfield 's classification, it is included here under the L2 proficiency construct, and assuming that integrating background knowledge (using higher level processes) with speech signals (using lower-level processes) to improve understanding is usually a trait of successful listeners. In the case of listening, Vandergrift (2007) suggests that during the process of listening, listeners build a conceptual framework to interpret the spoken message, using their prior knowledge (cultural background, familiar topics, believes, interpretation of non-verbal information, ideas stored in the long-term memory) as a base. This framework tends to place greater attention on information that match previous ideas and believes, while neglecting or losing new information that does not confirm the already existing ideas and knowledge about the world As a matter of fact, researchers have found that rich background information can help the listener to compensate for the lack of linguistic information, unclear speech, misunderstandings, or lack of some local information or context from earlier parts of the text (Bloomfield et al., 2010; Goh, 2000; Tayler, 2001).

Metacognitive strategies

Metacognitive strategies refer to the type of strategies used by listeners to direct and reflect on their own thinking while listening (Vandefrgrift and Goh, 2012). These types of strategies can help listeners to use more convenient and effective ways of listening in order to comprehend the message properly, avoid unsuccessful strategies that may lead to confusion or misunderstanding, and compensate for unknown vocabulary and coping with difficulties. Vandergrift and Tafaghodtari (2010) have developed a questionnaire that assesses listeners' use and awareness of metacognitive strategies (the metacognitive awareness listening questionnaire, MALQ). Accordingly, five factors are revealed as main metacognitive strategies that listeners may use:

- Planning and evaluation strategy: for preparing for listening and evaluating their comprehension level of a given text.
- Problem solving strategies: for making inferences and testing them using experience and general knowledge.
- Avoiding mental translation: skilled listeners try to avoid literal translation which is often a characteristic of low-level listeners.
- Self-knowledge: being aware of the difficulty level of a text and accordingly able to identify their confidence or anxiety level while listening.
- Directed attention: to stay alerted during listening and quickly recover attention when concentration fades away.

The use of metacognitive strategies is also recognised as an important component in the communicative language ability (CLA) framework proposed by Bachman and Palme (1996) as explained earlier and was discussed under the strategic competence component of that framework, although the areas of metacognitive strategies explored under the CLA framework are slightly different from the above ones suggested by Vandergrift and Tafaghodtari (2010). However, the core purpose is relatively the same. In the case of L2 listening tests supported with visual materials, it is believed that successful use of metacognitive strategies can help in constructing meaning from the two sources of information, audio and visual, by relating the incoming information from these sources with the linguist and non-linguistic knowledge of the listener, and also with the context of speech (Wagner, 2006b).

Anxiety

Anxiety can have a negative effect on the listener. If the listener is worried that the message is above his/her level of cognitive abilities and it is too complex, the ability to concentrate on the text falters and comprehension decreases (Bloomfield et al., 2010). However, the level of anxiety may possibly be reduced by some factors like note taking (Chang and Read, 2008) and assisting understanding by providing some cues, like nonverbal or visual cues (Ginther, 2002; Suvorov, 2013; Wagner, 2006a).

Characteristics of the text

There are many characteristics that can be related to the spoken text and affect its level of difficulty of L2 listening text. The most significant factors that will be covered in this review are authenticity, text length and complexity, organization, and auditory features.

Authenticity

Authenticity of the text is one of the important characteristics that educators have emphasized in language tests (Bachman and Palmer, 1996). The term authenticity is defined by several researchers according to different factors like authenticity of the passage itself or authenticity of the task associated with the passage, as explained earlier in section (2.3.3).

Pedagogically, it is advised that the use of authentic passages in listening as opposed to passages created specifically for listening can potentially provide listeners with many advantages and benefits (Kienbaum, Russell, and Welty, 1986). For instance, regular listening to authentic texts allows the listener to gain experience with elements of the target language use like phrasal verbs and lexical reduction forms (Flowerdew and Miller, 1997). In the case of listening assessment, a number of researchers argue that the level of authenticity of a text can be increased by providing visual materials (Suvorov, 2013; Wagner, 2006b, 2007, 2010), as it is relevant for the current study.

Text length and complexity

Text length is one of the key factors that affect difficulty, and due to its importance, it has been measured in several ways in the literature. For instance, different techniques employed for this purpose including counting the number of words and sentences or measuring the duration in minutes. Normally, longer listening texts contain larger amount of information (Bejar et al., 2000; Rost, 2006). This implies that lengthy passages may negatively affect listeners' comprehension as they impose heavier load on working memory (Henning, 1990). However, inspecting only the length of a passage, in isolation from other factors, cannot reveal a reliable correlation, and other factors should be taken into account. Among these factors according to Bloomfield et al., (2010) are information density and redundancy:

- Information density: refers to the amount of information in a passage, which is regarded as more predictive element for difficulty level than the text length. A piece of information has been defined in the literature in different way, some researchers (Nissan, DeVincenzi, and Tang, 1996) regarded *content words* (e.g., verb, noun, adjective, etc.) as a sole measure of the amount of information per passage, and the procedure followed was counting the content words and dividing it by the total number of text words. They actually found no relationship between text difficulty level and the amount of content words. Others (Rupp, Garcia, and Jamieson, 2001) considered another measure of content words which is the *type/token ratio* (number of unique words that are not from the same word family, e. g., cat and cats are from the same word family). This measure

  was a significant predictor for item difficulty as the larger use of type/token ratios increases the difficulty of the item. However, despite these measures stated in the literature, the reference to *information* in this study is a broader one that includes not only the linguistic/aural information, but also the non-linguistic/visual information which is naturally a part of academic texts. Hence, during the creation of visuals in the current study, the researcher intended to accompany the aural information with the visual by designing Power Point slides that reflect basic information stated by the speaker.

- Redundancy: refers to the presentation of information more than once by the use of repetition, elaboration, paraphrasing, or other methods. Field (2008) explains that speakers tend to go back to former points either to check for the listeners' understanding in case of interactional listening, or to stress the importance of a previously mentioned point. This factor can have a positive effect on the listener by providing a second chance to revise and understand the information presented in the text. However, this effect depends on the listeners' proficiency level and on the type of redundancy employed (i.e., whether repeating the same words or paraphrase them). Chiang and Dunkel (1992) found that higher ability students benefit from redundancy more than lower ability students. Simple repetition of the same words is regarded as the simplest form of redundancy, while paraphrasing and giving synonyms are more difficult and less salient (Kostin, 2004). In the current study redundancy is considered as a characteristic of academic texts, and therefore when recording the videos for this study, the researcher advised the speakers who participated in giving the lecture to feel free to repeat any word or explain words by giving synonym in a way that they would naturally do in their own real lectures. The researchers intended not to cut these parts in the process of editing the videos.

Organization

Another feature of the text that can affect listeners' comprehension is text organization. Generally, this feature refers to the degree of coherence in a text, the discourse markers and structure, and the position of information needed to answer test questions.

Coherence is difficult to define and measure. Generally, it can be described as logicality of a text, or how ideas are presented throughout the text (Freedle and Kostin, 1999). However, since there is a debate around defining and measuring coherence objectively; research results about its effect on listening comprehension are mixed. In general, the available results indicate that coherence has an effect on overall comprehension (Kostin, 2004; Ying-hui, 2006). Ying-hui (2006) states that a well-organized text would enhance comprehension because the ideas presented in the text are interlinked and logically presented. Buck (2001) explains that in lecture comprehension, listeners can benefit more from the overall inter-relatedness of the text than from understanding individual sentences.

Discourse markers are also a component of organization that may influence comprehension of spoken texts. These markers refer to the words and phrases that link the text together (e. g., ordinals: *first*; specifiers: *the only*; conjunctions: *yet*, *but*). Some researchers have found that different discourse markers have different effects on comprehension (Kostin, 2004; Ying-hui, 2006). For instance, there is evidence that discourse markers which organize the relationship between the overall structures of a text can help L2 listeners to comprehend, while markers that establish a relationship in a smaller level, which is between words and phrases, do not seem effectively helpful.

Another factor in organization that could be considered here is position of information that is important to answer test question (e. g., at the beginning, middle, or end of the text). In fact, little research has been done on this area and it suggests that listeners can recall information that occurs towards the beginning or towards the end of a text more easily than information positioned in the middle of the text (Freedle and Kostin, 1999). Kostin (2004) maintained that test items which their answers come from information at the end of a text tend to be easier than other items.

Audio-related features

Other factors affecting listening comprehension are related to auditory features, like the noise or distortion level in the recording. Normally, these can affect comprehension even with L1 listeners and proficient L2 listeners (Adank, Evans, Stuart-Smith, and Scott, 2009). The most distracting noise is one that closely resembles the speech signal like babbling sounds (Bloomfield et al., 2010). Other sounds from other different sources, like audio signals or the noise of filtering out high frequency information, can affect listeners differently according to other various variables like their aptitude, motivation, and their anxiety level. This factor was controlled in this study by recording the videos in class rooms or auditoriums with no audience except of the speaker, the researcher, and a camera operator. In addition, the sound was filtered in the process of editing in order to ensure that the sound is of high quality. Another procedure to ensure the quality of sound in this study is using headphones instead of playing the sound through the speaker. Somers (2007) claims that there is evidence that test takers perform better when using headphones (cited in Elliott and Wilson, 2013) because noise from the surrounding context can be cut.

Characteristics of the speaker

The difficulty of a spoken text can be directly affected by auditory factors other than the noise level, most importantly factors related to the speakers themselves. Some researchers have included this section as a component of the auditory features of the text explained in the previous section (e.g., Bloomfield et al., 2010). In the current study the speakers' characteristics are discussed separately, as the researcher considers it to be of a great importance.

It can be said that familiarity with the speakers' accent is one of the important factors that can affect L2 listeners as well as L1 listeners, though to a lower degree in the latter case (Major, Fitzmaurice, Bunta, and Balasubramanian, 2005). Familiar accents can be easier to understand than unfamiliar ones (Bloomfield et al., 2010), which may interrupt comprehension even if the text itself is not difficult in terms of its ideas, information, and syntax.

Some other features related to the speech are disfluency and speed. Disfluency refers to some acts that could, to some degree, affect the flow of speech like hesitation, pauses, fillers, or repetition of some words. Many researchers have inspected this factor (Arnold, Fagnano, and Tanenhaus, 2003, 2004; Collard, Corley, MacGregor, and Donaldson, 2008; Watanabe, Hirose, Den, and Minematsu, 2008) and most of these studies suggest that these features can aid and support comprehension. For instance, pauses and hesitation may give the listener additional processing time to figure out the meaning of what has been said, especially with higher level learners.

The speech rate or speed is also a significant factor that could possibly have an impact on comprehension. Several studies have revealed that high speed speech can have a negative effect on comprehension for both beginners and advanced listeners (Griffiths, 1990, 1992; Rosenhouse, Haik, and Kishon-Rabin, 2006). Interestingly, it has been also found that while listeners are negatively affected by high speech rates, they do not find low speech rates to be beneficial to their comprehension (Derwing and Munro, 2001). This implies that when choosing a speaker to present a talk or a lecture in a listening test, their speech rate should be normal.

In the current study, speakers with English and American accents are used. All of the speakers are characterised by their clear articulation. The speakers were advised by the researcher to act naturally in terms of accidental disfluencies acts like falls start or hesitations. Also, as they are all experienced lecturers, their speech rates are moderate in order to be convenient for international students' needs.

Characteristics of the tasks and test condition

This factor in L2 listening comprehension tests lies completely under the control of the test designer. Starting with test tasks, test developers can choose from a variety of tasks according to their test construct. For instance, some listening tasks may require the test taker to recall specific and detailed parts from the passage. Other tasks may ask about comprehension of main idea in the text. Therefore, the choice of question types will differ according to the construct applied, so for example in some cases multiple-choice questions will be used, in others open ended questions. These different choices of tasks would most definitely affect different proficiency listeners to different degrees. However, Field (2013) asserts that a balanced variety of listening task types should preferably be provided to test takers in order to guarantee that no specific cognitive style is privileged over the others.

Other factors are related to the testing condition. Most importantly: time limits, note taking, and repetition of the passage. All these factors can affect comprehension and the level of difficulty of the test to certain degrees.

- Time limits: normally, any task that is timed would seem more difficult. In the case of the effects of time limits in L2 listening comprehension tests, little research has been done to date. However, research on cognitive tasks and performance of examinees in timed conditions revealed that the limited-time variable can negatively affects test takers' working memory capacity (Siemer and Reisenzen, 1998). Time-limits in listening comprehension is mainly related to the task time, as time for listening to the text occurs on-line and can be only increased by repeating the passage. In this respect, Buck (2001) argues that increasing the task time will not improve test takers’ performance, since they cannot refer back to the text if they did not understand it the first time. Therefore, in this study, in order to control the overall time of the test, the researcher decided to allow the participants to view the items (i. e., five items after each text) for 2 minutes maximum.

- Note taking: most research evidence available in the literature suggests that note taking can be beneficial to test takers (Carrell, Dunkel, and Mollaun, 2002; Lin, 2006). However, an exceptional study done by Hale and Courtney (1994) argues that note taking is not helpful and test takers’ performance is better when they do not take notes while listening because it imposes an additional effort on the listener which may lead to confusion. As a matter of fact, this result seems to over-generalise the effects of note taking. Lin (2006) stresses the importance of not forcing the participants to take notes while listening and give them the freedom of choice whether to take notes during listening to the text, which is the case for this study as participants will be provided with a piece of paper and a pencil in case they want to take any note. It has also been suggested that optional note taking; even if it is not significantly advantageous, would not be detrimental to the listeners (Bloomfield et al., 2010; Lin, 2006).
- Double hearing: some researchers have examined the effects of multiple hearings (or double hearing) on the performance of test takers and suggested that replaying the text can support and enhance comprehension, especially for lower-level listeners (Field, 2008; Jones, Pearson, and Glyn, 2011). However, the relation of the possible advantageous effect of multiple hearings and proficiency level is not consistently reported across studies since different effects were observed with different listeners. In addition, some researchers have considered that repeating a passage can affect authenticity and should be avoided (Fortune, 2004; Suvorov, 2013). However, Jones (2011) suggested that this factor has not been explored in a conclusive way and still needs to be researched more accurately. For practicality and authenticity reasons, each text will be played only once in the current study.

These four characteristics (of the listener, of the text, of the speaker, and of the task and test condition) cover a range of the important factors that may generally affect performance in L2 listening tests, and in particular, these factors may have a direct impact on test takers' cognitive processing. As the current study exploring the effect of visual materials in L2 listening test, the cognitive processes will be explored from the angle of viewing patterns of test takers. In other words, the viewing patterns as revealed by the eye trackers are used as an indicative of the cognitive processes that test takers' may have employed during the test. Before going into more details about the processes of linking viewing patterns to the cognitive processes, it is important first to provide an outline of the use of visual materials, which switches attention to the following section on the role of visual materials in L2 listening tests.

2.4. Types of visual materials in L2 listening.

Visual information comprises an indispensable part of our communicative nature as human beings. In language classrooms, visuals have been incorporated as an important tool that supports instruction. However, its use in language testing still seems to be in its early, preliminary stages, lacking the needed empirical evidence to be fully embraced. As the core purpose of this study is exploring the impact of including visuals in L2 listening tests, it is essential to provide an outline of the types of visual materials, their use in L2 listening comprehension, and viewing the available body of research on the impact of including visual materials on test takers' performance in L2 listening comprehension tests.

2.4.1. Classifications of visual materials.

In the field of second and foreign language listening assessment research, there is a clear distinction between two types of visuals. These are known as context and content visuals (Bejar et. al., 2000; Ginther, 2002; Ockey, 2007).

In general, context visuals are used in listening texts to show either the image of participants or the settings of the scene, like using a picture of a lecturer in a conference room. More particularly, context visuals refer to the visual elements that are related to the environment in which the spoken stimulus is presented (Ginther, 2002). An example of this could be a series of photographs depicting the speaker and the setting. For instance, a professor giving a lecture in a classroom would be considered a context visual. According to Bejar et al., 2000, there are three sub-types of context visuals, based on the information they convey. The first type conveys information about the environment, such as a picture of a classroom. The second type provides information about the participants involved in the oral input, like a picture of the teacher or lecturer. The last type is visuals that contain information about the type of text, such as a visual of a student giving a presentation.

The second type of visuals, that is content visuals, is the one illustrates the content of the verbal information by using aids like still pictures, maps, diagrams, or a chart. For example, using a picture of plastic waste on a beach with a text about saving the sea life. As indicated in the study of Bejar et al. (2000), content visuals are classified into four types, each one plays a specific function related to the verbal input. These are: (a) visuals that replicate information in oral stimulus; (b) visuals that illustrate the oral stimulus; (c) visuals that organize information in the stimulus; (d) visuals that supplement the oral stimulus. However, Bejar et al. (2000) point out that while the first three types of content visuals have the potential to facilitate comprehension, the fourth type can possibly make the comprehension of audio texts more difficult.

Nevertheless, although this distinction has been used by a number of researchers, it was found that the practical application of this classification is somehow problematic, since there is an overlap between the two types, and even more overlap and lack of clarity among the sub-types of each. This overlap appears clearly in all content visuals, which in addition to providing the information about the content; they also convey context information, like the image of the speaker or the lecture room (Suvorov, 2013). This inadequate distinction led some scholars (e.g., Peterson, 2001) to argue that there is an element related to both types; content and context, in all visuals, therefore, they cannot be strictly distinguished as two separate types.

Suvorov (2013) has provided a new taxonomy of visuals (instead of the traditional content and context classification) which is based on the results of his study on the effects of content and context visuals on test takers' performance in listening comprehension tests. He stated that "the current classification should be refined and further developed to include other dimensions of visuals" (p. 216). The eclectic approach that he adopted resulted in presenting a multidimensional taxonomy of visuals; with each of the dimensions represent the visuals on a *continuum*, with no rigid boundaries between them. The first dimension is called *semantic congruity*, which refers to the congruity between the auditory and visual information. This asserts the fact that the use of video in listening tests should not be as a decorative device added to the test, instead, it has a more important role in facilitating comprehension and improving performance. Semantic congruity can be explained through five types of relations between the visual and verbal stimuli, as suggested by Shriver (1997): (1) Redundant, which is when the visuals reflect identical meaning to the verbal input. (2) Complementary, when visuals provide different content from the verbal input, and both modes are important for understanding the main idea of the text. (3) Supplementary, when visuals and verbal stimuli present different content but one of them provide the primary idea and the other supplement it. (4) Juxtapositional, when visual and verbal stimuli present discordant ideas semantically, but both need to be simultaneously presented in order to infer the main idea. (5) Stage-setting, it also refers to visuals and verbal stimuli that provide different content, but one of them present the main idea and the other present the content.

According to Suvorov (2013), semantic congruity can have a positive effect on test takers' performance, because when the visual information is congruent with the verbal stimuli, that can lead to better comprehension of the text. Suvorov, following the ideas of Hu and Jiang's (2011) cross- model study of semantic integration in L2 listening comprehension; has classified the semantic relationship between visual and verbal into three conditions: semantically congruent, semantically neutral, and semantically incongruent. It is sometimes difficult to know how exactly the visual information is related to the verbal input. However, in the current study, the researcher attempted to design the Power Point slides for each text in a semantically congruent way, that provide visual information that is congruent with the verbal input, as it will be explained in more details in the methodology chapter.

The second dimension, which is very significant, is called *dynamism and movement.* This dimension refers to the amount of temporal and spatial transformation or dynamism in the visual content. Conventionally, visuals have been categorized into two main types according to the mode of delivery: static visuals and dynamic visuals. The first refers to still pictures and images, and the second refers to videos. However, Suvorov (2013) suggests that movement and dynamism do not actually depend on the delivery mode as these would traditionally be associated with videos only. Instead, he explains that this dimension depends on the visual *content*. For example, videos that show a lecturer speaking in an auditorium while she/he is standing or only moving for several minutes are recognized as static because the content is the same. On the other hand, a set of content-based pictures that change in a particular sequence (e.g., every 15 seconds), with each picture deliver new visual content can be argued to be dynamic because the content changes with every picture. Therefore, in this dimension, Suvorov suggests that visuals can be classified into three main types: a) dynamic, b) static, c) liminal, all of which are based on the semantic content of the visual.

Hence, static refers to those visuals where there are no major changes in the spatial or temporal conditions, for instance, a video that shows the head of the speaker or a single picture of a map in geography lecture. Dynamic, on the other hand, refers to that type of visuals where the spatial and temporal conditions are continuously changing. This results in changing the semantic content as well. An example of this type is a set of pictures that show the stages of chemical experiment, or a video of a professor explaining the process of how an electronic device works. The last type of visuals in dynamism and movement dimension, liminal, can be viewed as falling somewhere in the middle, between the other two types (dynamic and static). This type represents a transitional phase between dynamic and static, by combining aspects of both. An example illustrates this type can be a student giving a power Point presentation, where the student is located in one side of the screen and the changing Power Point slides occupy the other side. These slides change at certain intervals, changing the content as well. Here, one can say that these visuals possess features of static visuals because the speaker is not changing in the temporal or spatial condition. In addition, this type of visuals is also dynamic because the semantic content of the Power Point presentation changes with every slide. Relating this to the visuals used in current study, it can be said that almost all visuals are liminal as they view the speaker beside a screen with changing slides. This will be discussed in more details in the methodology chapter.

The third dimension, according to Suvorov (2013), is the *rhetorical effectiveness* of visuals. Elements of visual rhetoric include graphic, spatial and textual aspects. This dimension refers to how visuals can deliver information in an accessible and persuasive way. Suvorov has found that certain elements in the visuals might attract test takers' attention and influence their responses, like the speakers' personality or his/her use of gestures and body language. Accordingly, visuals can be classified as: rhetorically effective, rhetorically neutral, and rhetorically ineffective. But the question is how visuals can be rhetorically effective in the first place? Kostelnick and Roberts (2011) suggest that visuals' rhetorical effectiveness can be based on six interdependent factors: (a) arrangement (e.g., the organization of visual elements), (b) emphasis (e.g., visuals that are more important and dominant than others) (c) consciousness (e.g., the degree of complexity and details of visuals), (d) clarity (e.g., how easily to understand visual elements), (e) tone (e.g., the degree of formality of visual elements), and finally (f) ethos (e.g., the degree of credibility of visual elements). According to these factors, visuals' rhetorical effectiveness is determined by the rhetorical situation, which is composed of the participants, the context, and the purpose of the visual text (Kostelnick and Roberts, 2011). Furthermore, it is assumed that the degree of rhetorical effectiveness of visuals can influence the viewers' attention and interests, and evoke different emotional responses (Kostelnick and Roberts, 2011), and it can also play a role in activating their background knowledge (Ockey, 2007). This means that this dimension can actually affect the way of perceiving visuals by the viewers and the impact that it may make on them. In other words, as rhetorical effectiveness of visuals may influence test takers' attention to the visual materials (e. g., where they look and for how long), it can be said then that it could have a direct impact on test takers' cognitive processes. With the use of eye tracking technology in this study, the overt attention can be detected, and by linking these levels of attention to test takers' verbalisations of their way of perceiving the visual materials and listening to the audio input simultaneously, a clearer indication of the cognitive processes employed might be obtained.

As these three dimensions (semantic congruity, dynamism and movement of content, rhetorical effectiveness) are not separated from each other; they can be considered as components of "visual language" as suggested by Petterssons (2002). This term refers to the structure and semantics that visuals can convey. Petterssons (2002) argues that visual language, like verbal language, is governed by meaning (semantics), structure (syntax), and content of use (pragmatics). Hence, relating this to the three dimensions of visuals explained previously, one can argue that visual language can present *meaning* that reflects the verbal message, which is identified by the semantic congruity dimension.

Visual language can also have a *structure* presenting the spatial and temporal conditions of the verbal content, which is identified by the dynamic and movement dimension. And lastly, elements of visual language in specific contexts can affect the way visual message is perceived *pragmatically*, which can be related to the rhetorical effectiveness dimension.

The multidimensional taxonomy of visuals as proposed by Suvorov (2013) is perceived in this study as an important factor in deciding and shaping the type of visuals to be used, since the content-context classification is not considered as an accurate distinction between types of visuals. Therefore, when designing the visual materials of the current study, it is decided that visuals in the current study are designed to be semantically congruent with the verbal message and liminal in its visual content. This decision was based on the L2 listening construct employed for the test in the study, which stresses the importance of assessing the ability of test takers to process and understand both acoustic and visual input from academic lectures in university settings. In other words, the theoretical definition of the construct places an equal weight on the importance of both types of input, visual and acoustic, and does not specify that one of these types of input is the dominant and the other is the supplement. The construct also refers to the setting as academic lectures in university setting. Therefore, when operationalizing this theoretical base of the construct, the design of the visual materials had to reflect a real academic lecture, where both the lecturer and the supporting materials (PPT) are visible to the listener. The visual content is chosen to be liminal purely because of practicality reasons related the use of eye tracker. The speakers had to appear in one side of the screen to make the analysis to eye tracking data systematic, as will be explained in the methodology chapter. However, the last dimension (rhetorical effectiveness) cannot be decided by the researcher, as this can only be perceived by the viewers of the visual listening test, whether they find it rhetorically effective, neutral, or ineffective. This dimension will be achieved by asking participant during the cued retrospective reports about their perceptions on the two types of visual materials.

2.4.2. The role of visual materials in L2 listening comprehension

As mentioned earlier, the impact of visual materials on L2 listening comprehension is still controversial. According to some researchers, incorporating visual materials into L2 listening may enhance the performance of L2 learners. (Ginther, 2002; Gruba, 1997;

Ockey, 2007; Rost, 2013; Suvorov, 2015; Wagner 2013). Others argue that visual materials in L2 listening should not be included because they represent an irrelevant variable to the purpose of listening (Buck, 2001; Coniam, 2002). In fact, several studies in L2 listening comprehension field revealed that visuals can help L2 learners to develop their listening skill (Baltova, 1994; Wagner, 2006a, 2010) in a similar way to what has been found in L1 language acquisition where children are found to rely on non-verbal information to process vrbal messages (Safarali and Hamidi, 2012). However, it seems that these studies did not present enough evidence to end the state of debate among researchers. In the literature, many researchers have attempted to highlight the strengths as well as the weaknesses of including visual materials as part of listening teaching and assessment.

The effect of visuals that accompany texts (written or spoken) has been explained by a number of theories. Dual coding theory by Paivio (1991) is usually referred to in the literature as one of the earliest theories that dealt not only with text comprehension, but also with the effects of visual display in comprehending messages in reading and listening. The theory aimed to give equivalent weight to verbal and non-verbal processing. According to this theory, visual information and verbal information are processed through separate, but interconnected, cognitive sub-systems, which are known as the verbal system and the imagery system. Paivio argues that verbal messages (words and sentences) are processed only through the verbal system, while pictures can be processed through both the imagery and the verbal systems (Paivio, 1991). Thus, the high memory for visual information and the enhancing effect that pictures in texts have on memory is attributed to the advantage of dual coding as compared to single coding in memory. He also assumes that both verbal and imagery information can be kept in the working memory, which enables the individual to make cross-connections between the verbal and visual codes and that makes retrieving this information easier (Clark and Paivio, 1991). However, critics of this theory (Baddeley, 1992; Mayer, 1997; Schnotz, 2005) argue that the dual coding theory does not form a comprehensive and satisfactory base for understanding the way verbal and visual information is processed. It generally assumes that the addition of pictures to any type of texts has a positive effect on learning and neglects the fact that pictures can sometimes be detrimental if they interfere with the mental modal constructed through the verbal input. As a result, other theories evolved in order to provide more inclusive explanation to the nature of picture and text comprehension.

Among the different theories that attempts to explain the inclusion of visuals in listening texts, there are two major theories that presented more detailed accounts in this term. These theories are Mayer’s (1997) cognitive theory of multimedia learning, and Schnotz’s (2005) integrated model of text and picture comprehension. Starting with Mayer (1997) who hypothesized that learning can be facilitated through the idea of “multimedia instructional message” (p. 32), which is based on how the human brain works by integrating information from two sources: the aural channel and the visual channel. What is meant by multimedia instructional message is any form of communication for learning using words and picture. Words here refer to written words or spoken words. Pictures refer to static pictures, like photos, diagrams, and illustrations, or dynamic like in video clips and animation. This theory is basically based on three cognitive assumptions: 1) the dual channel assumption, confirms that learners process incoming information through two separate channels which are the auditory channel for verbal information and the visual channel for pictorial information, which is similar to this point to the assumption of dual coding theory. 2) The limited capacity assumption refers to the limited capacity of each channel to process given information at a time. 3) The active processing assumption indicates that learning occurs when a set of cognitive processes are integrated, that complement and confirm each other. In a series of experiments Mayer and colleague found that participants’ comprehension improves with visuals only in certain instances (not beneficial in all cases as the dual coding theory suggests). Learners’ comprehension improves when texts and pictures are explanatory, when the contents of visual and verbal input are interrelated, when visual and verbal material are presented closely together in space or time and also when the recipient has little previous knowledge about the subject area but high spatial cognitive abilities (i. e., the ability to understand and remember the relationship between objects) (Mayer, 1997).

Mayer (2005) argues that since verbal and visual information are processed by two separate channels, this processing results in building two parallel mental models. Subsequently, these models will be integrated in the cognitive system and represent each other by making connections between the text-based model and the picture-based model.

However, integrating the two models can occur only when the verbal and visual information are available simultaneously in the working memory (Baddeley, 1992).

As texts and visuals are grounded on different sign systems and use quite different principles of representation, the idea of parallelism of visual and verbal processing that Mayer’s theory suggested seems problematic to some degree. That is because these different representations have different uses for different purposes (Schnotz and Bannert, 2003). More precisely, texts are more powerful in discussing different types of ideas and the relationship between them, while visuals are more powerful in drawing inferences (Johnson, Laird and Byrne, 1991). Therefore, Schnotz and Bannert (2003) have suggested what is known as the integrated model of text and picture comprehension as an alternative model to Mayer’s cognitive model of multimedia learning. This model consists of two main branches: descriptive and depictive branches of representations. The descriptive branch is mainly related to the use of *symbols* to represent meaning. These symbols are signs which are associated to their referent by convention. For instance, the word tree has nothing to do with a real tree, but it is related to its referent by convention. The processing of symbols results in an interaction which comprises the given (external) text and its (internal) mental representation and semantic content. On the other hand, the depictive branch uses *icons* which are also signs, but related to their referent by similarity, like a picture of a tree (Schnotz, 2005). The processing of icons results in an interaction comprises the given (external) picture and its (internal) visual perception, in addition to the internal mental representation to the subject matter of the picture presented in order to understand what the picture means instead of merely perceiving it as it is. Hence, mental models are not constrained to a specific sensory model (the eye or the ear). Instead, mental models are constructed via auditory perception as well as visual, haptic, and kinesthetic perception (Schnotz, 2002), and stand on working memory (Baddeley, 1992).

The above theoretical account provides an attempt to understand and highlight the important role that visuals play in comprehending verbal information (written and spoken). However, it is not completely clear yet how pictorial information together with the textual information can be integrated and in what way. New theoretical frameworks (e. g., Verhoeven & Perfetti, 2008) highlight the developments in understanding the processing of sources of information as cross-modal and that involves communication in various directions between verbal and visual input. The directions cannot be accurately specified as they depend on the particular situation where the language interaction happened, but what can be stated at this point is that processing multimedia information is not linear. In other words, language users invest in different sources of information simultaneously. These sources can be textual, situational, or rooted in their prior knowledge (Zwaan, Kaup, Stanfield, & Madden, 2001).

However, it is not yet clearly understood how visuals can affect the performance of L2 listening test takers, contrary to the abundance of empirical evidence available that supports the use of visuals in teaching listening comprehension. Research results on L2 listening assessment are still largely inconsistent as shown in the following section.

2.4.3 Previous research on visualised L2 listening tests.

There is a large body of research in support of using visualized texts in language classrooms (Chung, 1994; Mueller, 1980; Rubin, 1990, 1994; Secules et al., 1992; Zhou and Yang, 2004), and that resulted in the practical and successful employment of visuals in L2 listening comprehension classes. Yet, research on the effects of using visual texts in the assessment of listening ability is still in its early stages, and it is hotly debated as the results are greatly conflicting and inconsistent (Balatova, 1994; Ginther, 2002; Gruba, 1997; Ockey, 2007; Progosh, 1996; Suvorov, 2009, 2013; Wagner, 2006b, 2007). This, as discussed earlier, can be attributed to the lack of a widely accepted definition of the L2 listening ability, no doubt because of the numerous processes and variables that listening involves which make defining this ability a difficult task. Moreover, it is well established in the literature that the different effects of visuals on the performance of test takers in L2 listening comprehension tests could also be attributed to the different types of visual stimuli, texts, tasks, and contexts that are used in different listening studies (Ginther, 2002; Moore and Dwyer, 1994). The attempt to interpret the possible interactions among all these factors results in an endless process of analysis. Cronbach (1975) has provided an interesting description of this process as entering into “a hall of mirrors that extends to infinity” (cited in Ginther, 2002, p. 138).

The available literature on the field of listening assessment gives a flavour of the current debate on whether visual materials in general should be incorporated in listening tests or they should be abandoned. Ginther (2002) for instance, examined the effect of visual materials in the form of still images (present or absent), type of stimuli (dialogues /short conversations, mini talks, and academic discussions), and proficiency level (high or low) on performance of 160 students on the TOEFL computer-based test (CBT) of listening comprehension. According to Ginther, the most important reason for including visuals in CBT is that:

> Item stimuli including visual accompaniments to the audio text are considered better representations of actual communicative situations, so the inclusion of visuals may enhance the measurement of the test taker's listening comprehension. (p. 134)

The types of visual used were context and content visuals. The context visuals were mainly used to provide information for two purposes: 1] settings the scene in which the verbal exchange happens (i.e., a photo of classroom, department reception, etc.); and 2] revealing whether the speakers change in a conversation (i.e., a photo of two girls talking to each other). Content visuals on the other hand show and depict key contents of the verbal interaction, like showing maps or diagrams (i.e., a photo of meteor). However, Ginther used content visuals with mini talks only. She used the four types of listening comprehension items that are used in the TOEFL CBT item pool, which are accompanied by visuals. All tasks were presented as multiple-choice questions, which depend on the comprehension of specific details and the gist, either these were presented explicitly or implicitly. These types are as follow:

1) Dialogues/Short Conversations with context visuals (a still photo of the speakers)

2) Academic Discussions with context visuals (still photos of the speakers)

3) Mini talks with context visuals (still photos of the speaker)

4) Mini talks with content visuals (photos, diagrams, or drawings related to and contiguously presented with the content of the audio portion of the stimulus). (Ginther, 2002, p. 148)

All of these types were further divided into two sub-sets, with and without visuals. The results of this study indicated that content still images in mini talks were found to be facilitative, for both high and low proficiency level learners compared to participants who listened to the same texts with no visual support, when it provides information that complement the audio. Context still images, on the other hand, yielded different effects, with a debilitating effect in mini talks, no significant effect in dialogues/short conversations, and a facilitative effect in academic lectures.

Wagner (2006b) focused only on the use of videos in listening tests and suggested that video texts can play a positive role on test takers' performance, and that video texts can facilitate comprehension through the use of a non-verbal component. In his study, Wagner used two different types of stimuli: dialogues and lectures, and a total of six tasks, three for dialogues and three for lectures. Limited response (short answer) and multiple-choice questions were used, depending on the comprehension of specific details, general ideas, and inferred information. These items mainly put the spotlight on explicitly as well as implicitly stated information. The test was presented to two groups of test takers, one with the presence of video and the other with audio-only format. The participants in this study were from different linguistic and cultural backgrounds, and from intermediate and advanced levels. Test takers were videotaped during the test, using a Sony digital camera positioned next to the video monitor in order to record their viewing behaviour. The test was followed by a verbal report, by which participants can speak aloud what they are thinking while they were watching and answering the video text in the test. The results revealed that the video group had outperformed the audio-only group, in all components of the test (dialogue/ lecture texts, explicit/ implicit information, MCQ/ short answer questions). This means, as suggested by Wagner, that there were no test method effect contributing to the video-group’s superior performance. Wagner argues that the inclusion of videos in listening tests could be greatly advantageous, since it simulates the authentic spoken input and allow the participants to view the non-verbal behaviour, and therefore lead to more construct relevant variance in assessment. Consequently, the obtained results could lead to more valid inferences about test takers level, and how can they perform in real live situations.

However, Wagner’s (2006b) study has large limitations in the practical procedure that was followed to inspect participants viewing behaviour. When recording the participants’ viewing behaviour, Wagner used a camera, which did not cover the entire room where the test was administered. Some of the participants were sitting directly behind other participants, so they were not clearly recorded and as a result were excluded from the analysis. Analysis of data was done using a stopwatch timer. Wagner watched every participant individually in the videotape and started the timer whenever they made eye contact with the screen and stopped it when they move their eyes away. Obviously, this technique is not accurate enough (may be due to the lack of appropriate technology at that time) and the data obtained could not be regarded as precise data since it only gives information about how long the participants look to the screen but does not give any information about what exactly they are looking at, and whether they are attending to the non-verbal component in the video text or not.

Sueyoshi and Hardison (2005) investigated the impact of facial clues in listening comprehension on Korean and Japanese learners of English. Their goal was to consider if access to visual cues like lip movements and gestures facilitate ESL-students’ listening comprehension. Sueyoshi and Hardison randomly assigned three groups of participants, consisting of low-intermediate and advanced learners who listened to the same lecture, each in one of three different conditions: 1) audio-only, 2) audio-visual showing face only, no gestures and 3) audio-visual including gesture and facial expressions. The text used in this study was a lecture recorded specifically for the study. The results of this study revealed that the highest scores were achieved by participants in the audio-visuals with gestures and facial clues group, whereas the audio-only group scored the lowest. This implies that the effect of visual clues was positive, regardless of the participants' proficiency level.

Ockey (2007) conducted research to find out how test takers engage with different types of visuals that are still images and video in a computer-based test (CBT). The texts were taken from one lecture, where part of it was video recorded and another part was presented as a sequence of still images. Both texts were chosen from the same lecture in order to maintain the same topic and level of difficulty. The participants were international university level students. Ockey in his study employed observation, retrospective report, and interviews in order to determine how each of the test takers engaged with the two types of visuals. The results of his study suggest that participants engaged at minimal levels with still images, while they largely differed in their engagement with the video, ranging from very high levels of engagement to rather low or no engagement at all. Implications of that study were that video texts can be included in CBT, with a modification of the construct of listening of the test to include a reference to the *cognitive environment* which displays the natural body language, facial expressions, and gestures.

Recently, Suvorov (2013) has undertaken a study investigating how students interact with visuals in listening comprehension tests. In his study, Suvorov investigated the use of context video and content video, and the extent to which these two types of videos could affect test takers' performance and their interaction with the video. He applied the definition of context and content videos used by Ginther (2002). The six videos in this study were taken from real university lectures, three of them classified as context videos showing only the speakers and the setting around him/her (e. g., the classroom or auditorium in which the lecture took place). The other three videos were classified as content, showing some visual details about the content of the lecture (e. g., slides with textual clues, diagrams, or pictures of the subject that the speaker is talking about). An eye-tracker was used in order to reveal the viewing behaviour of the speaker and to detect any differences of the patterns of watching content and context videos among participants. The test was followed by verbal reports, in order to understand how test takers have perceived the two types of visuals. Results revealed that there is no significant difference in test takers' performance in the two types. Yet, the eye- tracker and the verbal report revealed that participants viewed content video significantly more than context video, and they also perceived it as more supportive and helpful. Suvorov implies that these results stress the importance of including visuals, especially content, so that the test will reflect the target language use situation more authentically and accurately. However, Suvorov's (2013) study covered two types of video texts and investigated the difference in listeners' perceptions and viewing patterns between these texts. It did not tackle the deeper level of investigating the cognitive processes that might be triggered by these video texts, which is considered a gap in the literature which the current study attempts to target.

On the other hand, other researchers have found that the inclusion of visuals in listening tests has a negative effect on the performance of participants. For instance, Gruba (1993) conducted a study in which he compared the performance of 91 intermediate-level students on 14 MCQ on an academic lecture. Two versions of the same lecture were presented, audio-only and video. The results revealed no statistically significant difference between the participant's performance in the two tests.

In a case study by Coniam (2002), he administered two versions of the listening test, audio-only and video, of the same listening test to two groups of non-native English language Pre-service and in-service teachers in Hong Kong. The text used was a discussion of a hotly debated educational issue. The choice of this text was based on a decision by the researcher supported by a testing committee as a text appropriate for the targeted audience. The test was followed by a questionnaire to elicit the participants' view of the two modes, video, and audio, of the test. Results of the test indicated no significant difference between the scores of the two groups. Furthermore, answers to questionnaire revealed that the participants in the video group felt no advantages for using the video, some of them even stated that watching the video distracted their attention. Therefore, Coniam suggested that listening comprehension tests should be delivered through the audio mode only, and videos should be excluded due to their distracting effect.

Suvorov (2009) conducted a study to investigate the effect of visuals on listening comprehension tests. The types of visuals were context single photograph and context video in addition to audio-only texts. Participants were non-native speakers of English from different nationalities and different levels of proficiency. The study was designed as a within-subject experiment, so all test takers went through the three parts of the listening test with: 1) a single photograph, 2) a video, 3) audio-only. Each of these parts included two texts, a dialogue, and a lecture. The test was followed by a questionnaire to ask the participants about their preferred type of visual stimuli. Results indicated that scores in the video part of the test were significantly lower than the two other types. In particular, the use of video with lecture text had a detrimental effect on students' performance. In addition, the test takers performance on their preferred part in the test, as indicated by their answers to the questionnaire, was not significantly higher than their performance on the other parts of the test.

When considering these studies, it seems that video is the most commonly used type of visuals in studies of L2 listening assessment. Despite this fact, language assessment based on videos is still considered an unexplored path because of the contradicting results these studies have yielded (Suvorov, 2013; Wagner, 2008, 2010). According to Jamieson (2005), the use of video in computer-based language tests is still avoided for three main reasons: 1] costs highly to produce. 2] needs advanced technological mastery to be produced and transmitted. 3] the effect of video on language construct is not clear yet. Gruba (2006) adds to these problems another one related to listening tests, which is the fact that the construct of video-mediated listening comprehension is not clearly defined and established so far.

It is noticeable that most existing research has focused on videos, by comparing video texts with audio-only texts (Coniam, 2002; Gruba, 1993; Sueyoshi and Hardison, 2005; Suvorov, 2013; Wagner, 2006b). Exceptions are Ginther (2002) in which she compared participants' scores on audio-only texts to scores on still images. Also, Ockey (2007) who compared students' engagement with video texts vs. still images. And Suvorov (2009) who compares participants' scores in all three situations, namely: video, single photograph, and audio-only. In spite of the fact that still images are already in use in some high-stake tests like TOEFL; but it seems that this use is not based completely on an empirical base. As Ginther (2002) stated that an important reason for using still images in CBT is "to enhance face validity" (p. 134), because presenting test takers with blank screen is considered inappropriate. Hence, the validity of using still images needs to be reconsidered, which is part of the current study's intention.

It is also clear that these studies have employed different procedures, different groups of participants (teachers, students), different text stimuli (dialogue, lectures, etc.), different item types (MCQ, short-answer, matching, etc.), and different text length and difficulty levels. All these varieties in test types and designs might have led to the inconsistent findings of research on visuals (Wagner, 2006). In addition, types of visuals have not been clearly reported in a number of studies, with exceptions to Ginther (2002) and Suvorov (2013), who clearly compared the use of content and context visuals.

The above account demonstrates that the effectiveness of different listening text conditions is not understood yet. There is no empirical evidence on the use of still photos in listening comprehension tests, in spite of being currently used in some CBT. The effectiveness of including video texts is also not known because of the different test circumstances used in different studies. Most studies used context videos only (or did not specify the type of visual at all) with few exceptions (Ginther, 2002; Suvorov, 2013) who compared context visuals with content visuals. This strict use of specific type of visuals ignores the fact that listening tests should in the first place reflect the language use domain if the scores to be used to make inferences about the test takers' actual level. In addition, test takers' perceptions on the use of different visual test conditions (video and still photos) is entirely absent. Hence, the current study intends mainly to fill in this gap by examining the same listening texts in three different conditions (Video, still photos, and audio-only), which are designed to reflect the natural academic situations which test takers are likely to encounter in their academic future. It will also investigate test takers' perceptions and how their performance is affected by these three conditions by employing a cued retrospective report.

Moreover, it is worth mentioning that the viewing behaviour of test takers to the video screen is virtually entirely ignored in these previous studies, except with Wagner's (2006b) study and Suvorov (2013). Wagner strongly argued that without exploring this aspect, we cannot attribute the difference in test takers performance, if there is any, on a video listening test compared to an audio-only test to the video, because we do not know whether they are actually watching the video or simply moving their eyes away from the screen. Empirical evidence has shown that attention spans are lowered gradually when watching videos used to teach foreign languages (Baltova, 1994). He states:

> …in the video conditions several students became distracted after six minutes, more students lost concentration after ten minutes and around one third of them kept watching until the end (p. 82)

However, as explained earlier, Wagner (2006b) used relatively old technology (digital camera) in order to record the participants' viewing behaviour, which can only give information about the amount of time test takers spent watching each video. It does not give any information about what items they are focusing on and for how long. Suvorov (2013), on the other hand, has used a more advance technology (eye-tracker) to investigate the viewing behaviour of test takers. However, the focus was to reveal the different viewing behaviour between context and content videos. Suvorov's study did not investigate the effects of non-verbal cues like body language and gestures on test takers' viewing behaviour, nor the cognitive processes that underlie their viewing behaviour. The current study intends, as a secondary purpose, to look for how the use of body language in visual texts, either with videos or still images, can affect test takers' processing and understanding of the text by using the eye tracking data and find out whether there is any link between test takers' perceptions and actual performance, and their viewing patterns.

Given that this study involves investigating test takers' performance and perceptions on the different conditions of visuals via the cued retrospective report, and also their viewing behaviour via the eye tracking, the following section is devoted to providing an overview of these techniques and their role in the study.

2.5. Approaches of exploring the role of visual materials in L2 listening tests.

In the field of investigating the impact of visual materials on the performance of test takers in L2 listening tests, researchers have mostly depended on comparative techniques, where they compare the scores of test takers in the visual and non-visual conditions (Ginther, 2002; Progosh, 1996; Wagner, 2010). As an alternative to this method, some researchers proposed the addition of qualitative techniques like retrospective and concurrent verbal reports and interviews (Ockey, 2007, Wagner, 2006b). Furthermore, investigating the viewing behaviour started to appear as a fundamental aspect of the process of investigating the impact of visual materials on the performance of test takers (Wagner, 2010; Suvorov, 2013). This section reviews the employment of these techniques with L2 listening tests.

More precisely, it explores the use of cued retrospective reports to extract test takers' perceptions and the use of eye tracking technology, which is used in the current study as a technique to record the viewing behaviour of the participants.

2.5.1. Cued retrospective reports.

Over the past two decades, researchers have attempted to develop a way to unravel the cognitive processes employed by language learners in order to understand how and why specific performances and information processing happen. One common way of doing so is to use verbal reports, which ask participants to report their thoughts and strategies for understanding texts and answering test tasks. This taps into participants' explicit knowledge of their behaviour, i.e., knowledge that can be accessed consciously. Traditionally, two types of verbal reports are used in research: concurrent and retrospective, where the first refers to reports that are extracted from test takers during task performance, and the latter refers to reports done immediately after completing a task (Camps, 2003). Both types are believed to be direct verbalizations of cognitive processes (Sasaki, 2003). In fact, each of these types has been widely used, and resulted in significant information that cannot be ignored. For instance, Whyte, Cormier, and Pickett-Hauber, (2010) have studied nurse performance and their associated cognitive processes.

In their study, they have applied both types of the traditional verbal reports (concurrent and retrospective). The results indicate that each of the two methods has contributed to reveal different areas and information about the inner cognitive process. The type of information revealed by concurrent reports is generally about actions and their consequences, whereas information revealed by retrospective reports refers to strategies for accomplishing a task and eliciting responses.

However, both protocols have some disadvantages and were criticized for not being sufficiently accurate (Schwarz, 2007). Some of these criticisms for instance, were that in the concurrent verbal reports, participants' responses may impede the ongoing cognitive processes. Retrospective reports are also criticized for being useful only in the case of very short tasks, while in longer tasks there is a risk of omitting information and thoughts that were present at some stage in the task performance and are forgotten during the report (Taylor and Dionne, 2000). For these reasons, researchers started to think about a new method to avoid the drawbacks of the two previous methods. The result was the *cued retrospective report* which is proposed by a group of researchers (Van Gog et al., 2005) and is based on a retrospective report, that is applied after completing the tasks, and is cued by a replay of the eye movement record of participants, or in some cases, a replay of mouse or keyboard operation (Sasaki, 2003). This method will avoid the interference of responses during the task, and it also avoids the risk of omitting information, due to the replay of the participants' record. Many researchers encouraged the use of this new approach of verbal reports, as for example Ehmke and Wilson (2007), who explained the employment of this approach, which “asks users to provide retrospective protocols cued by a replay of their eye tracking data to make it easier for them to explain their decisions and thoughts. This method is called PEEP (Post-Experience Eye-Tracked Protocol)” (p. 120). It is also explained further by Ball, Eger, Stevens, and Dodd (2006) who criticized the concurrent think-aloud protocols for being incomplete and can cause a number of complications to the participants because it forces them to verbalize ongoing cognitive processes during the performance of a specific task, which might be subconscious. This may lead participants to report what they believe they employed and cannot be guaranteed as a true indication of the real cognitive processes.

For this research, cued retrospective report are used to collect verbal information from participants about their use of non-verbal, visual information in the texts, and how this information could help them in answering test questions. The researcher used a set of guiding questions in order to extract information from participants as they verbalize their perceptions about the visualized texts. For instance, participants will be asked questions like:

- Which of the three types of texts (V, SP, A) did you find more helpful?
- While watching the video, what do you focus on? Why?
- While watching the still photos, what do you focus on? Why?
- Do you find the video/still photos helpful? If yes, what aspects are most helpful (i.e., speaker, face expressions, lip movement, gestures, etc.)? Why?
- Do you find the video/still photos distracting? If yes, what aspects are most distracting? Why?
- While answering your questions, did visual information help you to choose the answer?
- Place the three types of listening texts in the order of your preference. Specify reasons for your ranking order.

Linking cognitive processes to cued retrospective report data.

The verbal data that is collected from the cued retrospective report in this study is believed to be of a great importance in revealing how the test takers use the visual information provided in the visualized texts (video and still photo), and how they perceive this type of information. More precisely, the collected verbal data are used to seek evidence of cognitive processes used by the test takers during the test for processing the text and answering the test items. The cognitive processes that are of interest to the researcher in the current study are based on a cognitive validation framework proposed by Field (2012). This framework is established on comparing the processes that L2 listeners normally use in real-life, non-test situations with the processes they use under test conditions. The framework was designed to pursuit evidence of cognitive validity of L2 listening tests. Field (2013) defines Cognitive validity as “the extent to which a test requires a candidate to engage in cognitive processes that resemble or parallel those that would be employed in non-test circumstances (p. 78). As the core purpose of L2 listening test is to make inferences about the language ability of test takers in the specific TLU (target language use domain, in the case of the current study it is the academic domain/ lecture listening), it is thus important to find out whether the test employed can trigger cognitive processes that are normally used in that target domain. In order to do so, Field (2012) asserts the need to differentiate between three types of processes or behaviour, which are:

- Normal processes, which parallel the processes used by native listeners. Naturally, native listeners do not need to use specific strategies to inference meaning of a spoken text, instead they do that automatically and with little attention to specific individual words (Vandergrift, 2013). This type of processing is not of interest to the current study as the goal is not to assess native-like processes.
- Strategic behaviour, this in essence resembles the strategic competence component of the communicative language ability discussed earlier (section 2.1) by Bachman and Palmer (1996), which refers to the ability of L2 listener to prepare for and assess the task, and deal with comprehension problems by setting plans. Evidence of this type of behaviour reflects positive factor, as it means the test has manged to trigger cognitive processes similar to the ones used in real-life situations and that it reflects the test takers' language ability in the TLU.
- Test-specific behaviour, which refers to the specific strategies that a test taker might adopt to achieve a particular test task but would not normally use in non-test situation and does not reflect their language ability in the TLU, like guessing for multiple-choice questions. Evidence of this type of behaviour reflects negative factor, although it might not be possible to confirm that a test is completely against this behaviour (Elliot, 2013). However, excessive evidence of using test wise strategies is considered a challenge to the cognitive validity of the test (Field, 2013) because they do not reflect normal language abilities used in real-life situations.

Therefore, cued retrospective report data are basically used in this study to explore how test takers perceive visual information during processing the test and answering test questions. The verbal output by the test takers is believed to uncover their thinking processes during the listening texts supported by visual materials. In other words, the core purpose of the cued retrospective reports in the current study is to extract test takers' perceptions about the different text types used in the test, with a special focus on the video and still photo texts which are supported by the visual cues provided by the eye tracker. A secondary purpose of the reports is to explore evidence from test takers' verbalisations about their cognitive processes, specifically seeking evidence of their employment of two types of processes or behaviour as called by Field (2013), namely: strategic behaviour and test-wise behaviour. Evidence of the first is perceived as a positive behaviour that may support the validity of the test, while evidence of the latter is perceived as a negative behaviour that may weaken the validity of the test.

Besides the use of cued retrospective reports to reveal test takers' perceptions about the visualised texts and extract some insights about their cognitive processes during processing these texts and answering the test questions, this study attempts also to investigate test takers' interaction with the visual materials and explore their viewing patterns (with both video and still images) via the eye-tracking technology. Eye tracking is therefore used in this study for two purposes, 1) to generate heatmaps that can be shown to test takers, to better elicit their verbalisation and their perceptions about the video and still photo texts used in the test. 2) to collect data related to the viewing patterns of the participants in order to investigate their visual attention during the process of test taking. Below an overview of this method is presented, and a discussion of the cognitive processes that thought to underlie various eye-movement patterns is included.

2.5.2. Eye tracking.

"Eye tracking is the process of acquiring the spatial information corresponding to the movement of the eye as it unfolds in time" Reingold (2014). As a research tool, eye tracking is rapidly gaining popularity in many research fields, like cognitive psychology, sports, business, and education. The motivation behind recording human eye movements is basically to track the paths of visual attention which may give insights into the underlying cognitive processes used by the viewer. In other words, eye trackers usually used to calculate gaze movement in each task, that is how, where and in what order gaze is being directed. The reason for eye movement is basically the natural structure of the eye which, as explained by Carter & Luke (2020) restricts “high acuity vision” or the ability to see small details to a small segment of the visual field called the fovea. This restriction results in a continuous motivation to move the eyes to other segments where the fovea points at different stimulus that the person is processing or thinking about. This movement is commonly known as the eye-mind link (Just and Carpenter, 1984). It is claimed by many researchers that the eye-mind link aspect of eye trackers makes this technology an increasingly reliable tool for exploring and investigating questions relating to visual attention (Holmqvist, et. Al, (2011).

However, there has been some concerns that eye tracking technology by itself cannot provide accurate account of attention, as there are many other factors that affect the way people gaze at visual fields, where they look and for how long. These factors include many cognitive processes like language, memory, perception, and decision making. Hence, some researchers (Duchowski, 2007) suggest that the eye-mind link is not absolute and therefore eye tracking cannot be used solely to analyse attention. In other words, eye tracking data can suggest, for example, that the viewer has fixated his/her gaze on position X but cannot show whether their visual attention is on another area in the scene. For this reason, it is recommended that the use of eye tracker should be accompanied by other tools, such as verbal reports, in order to gain a complete and valid picture of attention, and extract what is beneath the overt movement of the eye (Scheiter and Van Gog, 2009). However, despite these concerns, it is generally true that the eyes movement do give insights into the mental processing of the items we are looking at (Holmqvist, et. Al, 2011). This can make eye tracking technology largely valid and applicable to many fields of research that studies areas of mental processes.

As mentioned earlier, test takers in this study are interviewed and asked to verbalize their perceptions and to provide reasons to explain their viewing patterns. They are given the opportunity to view their own viewing patterns via heat maps that are generated by the eye tracker as part of the cued retrospective report. In addition, the numerical data of eye tracking are also used to investigate the effects of implementing visuals in L2 listening tests on test takers’ performance. In other words, the researcher intended to examine the test takers’ viewing behaviours and to compare them to 1) their performance with the two visual text types (video and still photos), and 2) their perceptions about these two visual text types as elicited by the cued retrospective report. In this study, it is intended to examine the extent to which test takers watch and orient to the video and still photo texts, and investigate what aspects they focus on, and how they make eye-contact with the visual screen. Therefore, it is essential to review the use of eye tracking technology with listening comprehension, and how this technology can be used to link eye movements to listening comprehension.

Listening and eye tracking

When relating the term, visual attention, to the use of video in listening comprehension; it would refer to the directing of visual attentional resources to the video monitor. In other words, it is the amount of eye contact that test takers make towards the screen, and consequently their potential attention to the visual, non-verbal information provided within the text.

As a matter of fact, this aspect of investigating the test takers' attention to the visual stimuli in listening tests is not fully understood yet, since very little research had been done on this respect. As explained earlier in the previous studies section, the only two studies available so far that explored the viewing behaviour of test takers in L2 listening comprehension tests are Wagner's (2006b) and Suvorov's (2013) research. Only in Suvorov's study the eye tracker was used. However, it was used to explore the different viewing behaviour of test takers in context and content videos. Therefore, in the current research, as an attempt to make more accurate record of test takers' viewing behaviour in order to gain an improved construct validity evidence, the researcher uses eye tracker device to investigate participants' viewing patterns to video and still photo texts. This technology allows for recording test takers' eye movements during the test, as well as providing the cues for the retrospective report task. This will provide significant information about the patterns of watching both the video and still photo texts, amount of watching time spent with each text, and objects that have been attended to by the test takers (e.g., the speaker or the background scene).

The use of this device in listening tests is new, and so far to the best of the researcher's knowledge, there isn't any published study about the use of eye-tracking in visualized listening comprehension tests apart from Suvorov's (2013) study. Eye tracking is mostly used in research on computer-based tests with reading comprehension tests (Ashby, Yang, Evans, and Rayner, 2012; Bax, 2013) to investigate test takers' underlying cognitive processes that can be sought from their viewing behaviour during reading.

Furthermore, as explained earlier that because the eye tracking technology can only provide covert, implicit information about the participants’ viewing behaviour, that means it is not enough to rely on this type of data alone. This is therefore the value of combining this method with another sort of measurement such as verbal reports, which can provide significant insights into the inner cognitive processes in order to gain better understanding of participants’ performance in the test. This goal can be achieved by implementing a multidimensional research design, where different methodologies supplement each other. Hence, quantitative data gathered from test scores and eye tracking data, supplemented by qualitative data gathered from the cued retrospective report. The design will be explained with more details in the next chapter, methodology.

Eye tracking and cognitive processes.

As explained above, combining eye-tracking data with data collected from the retrospective reports is believed to provide a richer representation of the cognitive processes that test takers employ. For this reason, it is important in the first place to understand the measures that are used in eye-tracking.

Measures of eye-tracking:

Eye movements in general involve two basic types: fixations and saccades which happen either while watching scenes or displays or when reading any type of texts or information. fixation is commonly defined as a static state of eye movement, or the period during which the eyes are relatively fixed on some point of the visual field. Rayner (2009) explains that even though fixations are defined as stabilized eye movements, they are not completely static. Due to some peculiarities in the human visual system, fixations can sometimes involve micro saccades, drifts, or small shivers. Saccades on the other hand are simply defined as the rapid eye movement from one fixation point to the next. Many researchers claim that our visual input is completely suppressed during saccades, which means that we are actually blind when our eyes are performing saccades (Castet et al., 2002; Rolfs, 2015). According to Rayner (2009), the range of fixation periods differs from viewing one visual material to another. For instance, fixations generally range from 100 to 500 ms, and on average, they take about 250 ms while reading. Same applies to saccades. In reading tasks, saccades normally identified as a 2-degree and usually takes about 30 ms, while in scene perception tasks a 5-degree saccade usually happens, and it lasts about 40–50 ms. Most of the eye tracking measures that have been used in prior studies have focused on reading or word-based contexts.

Measurements of eye tracking have been identified through different approaches in the literature. For instance, Radach and Kennedy (2004) identified two major categories to measure word-based eye-tracking: the first category deals with the time of eye movement, (temporal scale), and the second category measures the space where eye movement takes (spatial scale). Using a different approach, Jacob and Karn (2003) listed six types of eye tracking measures, which they summarized as the most frequently used in tracking research. These measures are fixation count, average fixation duration, fixation count or rate, proportion of time spent on each area of interest (AOI), fixation count on each AOI, and gaze duration mean on each AOI.

AOI can be defined as a particular spatial area in the visual field that is specified in shape and position by the researcher who can during the analysis of eye tracking data define a chosen measure per AOI, such as number of fixations per AOI. Another tool that is useful for analysing data within a specific period of time is called Period of interest POI. This tool enables the researcher to analyse only the selected events that occur during that period and exclude events that occur before or after that period. For the current study, AOI and POI are both set precisely by the researcher to calculate three eye tracking measures, namely Fixation count, Dwell rate, and Total dwell time. More detailed explanation of these measures is provided in the methodology chapter.

The above section offered a description of some measures of eye-movements that can be recorded during various listening tasks; however, it is well established that different eye-movements are influenced by different factors and tapping into different underlying cognitive processes. The following section explains this aspect in more details.

Factors affecting eye movement.

As eye movement plays an important role in the behaviour of human beings, it can certainly be influenced by many factors (Rayner, 2009). Some of these factors are related to visual stimuli. For instance, when a visual stimulus or part of it is more eye-catching or complex, because of its bright colours or sharpness, it will attract more visual attention. Also, factors like the quality of the visual stimuli can affect the behaviour of eye movement. For example, dark or blurry stimuli need a longer viewing time, and bigger visual stimuli draw more fixations when compared to smaller ones (Henderson et al., 2014).

It has been noticed by many researchers that the movement of the eye can also be affected by cognitive factors. Just and Carpenter (1984) noticed that if the stimulus is not familiar, like an unfamiliar animal or abject, it will require a longer viewing time in order for the participant to recognise it. Similarly, it has been found that items that are less expected to be in a specific stimulus are viewed for a longer period of time (Henderson et al., 2014), while items that have an emotional content are processed quicker and attract more attention (Scott, 2009P).

These outcomes and observations have reinforced the need for more research to investigate how can eye movement give insights into cognitive processes. Researchers have made several attempts to find connections between measures of eye movement, that is known as eye tracking and inner cognitive processes (Boland, 2004; Viviani, 1990; Yarbus, 1967). For instance, one of the early attempts was made by the Russian scientist Yarbus (1967) who studied fixation patterns on pictures. The technique he used involved fixing a mirror system to the eyeball of the participant. He found that different questions about the same pictures provoked different viewing patterns as shown in Figure 2.2.

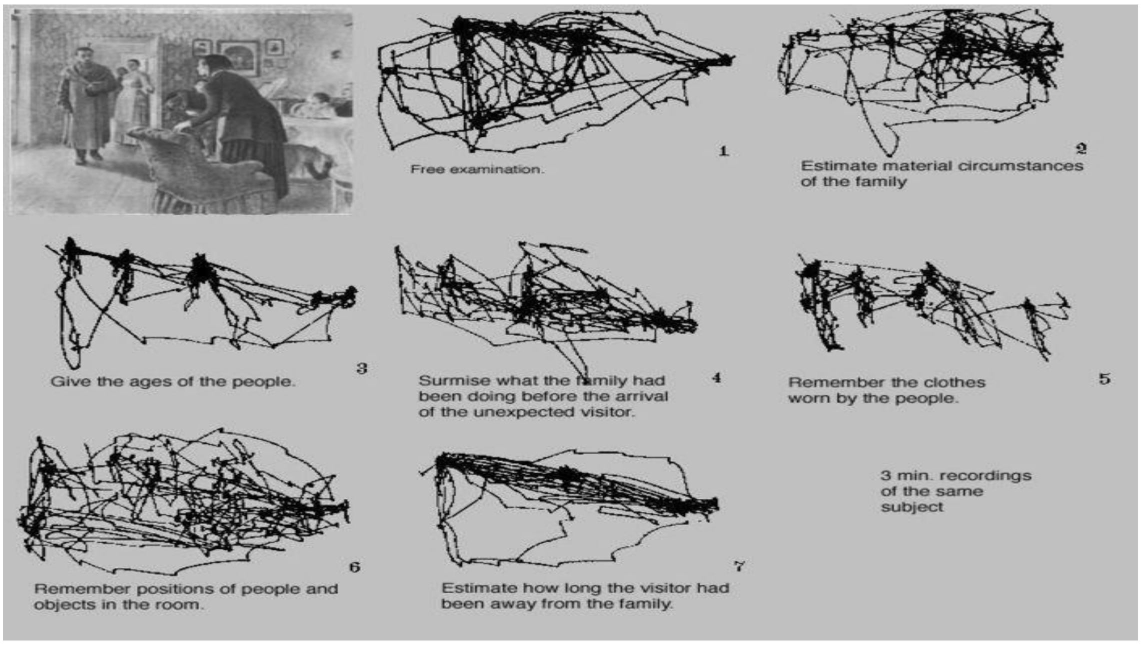

Free examination.
1
Estimate material circumstances of the family
2
Give the ages of the people.
3
Surmise what the family had been doing before the arrival of the unexpected visitor.
4
Remember the clothes worn by the people.
5
Remember positions of people and objects in the room.
6
Estimate how long the visitor had been away from the family.
7
3 min. recordings of the same subject

Figure 2.2: Different viewing patterns of a picture (unexpected visitor) with different questions. Yarbus (1967, p. 172).

Yarbus (1976) explains that these results show how the patterns of the eye movements of the participant were not simply drawn to noticeable or attractive parts of the picture such as bright areas. Instead, the fixation patterns reveal the elements that provide the most helpful information for each task are the ones that are being fixated the most. As the task changes, so does the viewer's eye movement to seek informative areas according to the new task.

However, it has been suggested that the use of eye tracking to explore the underlying cognitive processes may hold some problems. As discussed earlier, it is possible that viewers can sometimes pay attention to other objects around the point of their foveal gaze (Posner et al., 1980). For example, the overt attention of the eye is on a specific object in the visual scene, while the cognitive attention is processing oral input that does not relate to that object. Nevertheless, despite flagging this as a potential problem, many researchers assert that the viewing behaviour can be directly linked to the inner cognitive processes (Boland, 2004; Rayner, 2009; Suvorov, 2013; Yarbus, 1967). A well-known theory in this respect is the eye-mind hypothesis proposed by Just and Carpenter (1984), who claimed that there is a strong correlation between the viewing positions and the cognitive processes that are going in the mind. In other words, what a person is looking at is, to a great extent, what he thinks about. Despite the strong claims made against this assumption that the eye says nothing about the cognitive processes in the mind (Anderson, Bothell & Douglass, 2004), it can be still hold true in some cases and cannot be rejected completely. This hypothesis seems to present a relative rather than absolute case.

With respect to the current study, the choice of eye tracking measures is based on their perceived importance of reflecting some underlying processes. More precisely, the first eye tracking measure used, *Fixation count*, which is believed to be the most utilized parameter in eye tracking research (Jacob and Karn, 2003), is employed in this study because of its unique properties of indicating semantic importance toward the object fixated at. That is, more fixations on a particular AOI may indicate that the AOI is more important and/or more noticeable to the participant than the others (Poole and Phillips 2005; Holmiqvist et al., 2011). Therefore, it is generally perceived that the more important a given AOI is to the viewers, the more fixations they are probably going to land on the area. The second measure, *Dwell rate*, is commonly used to assess a participant's level of interest in an object or the amount of information it provides. However, a longer dwell time may also suggest uncertainty, lower situational awareness, and difficulty in extracting information from a display (ndrychowicz-Trojanowska, 2018). And the third measure, *Total dwell time* has a semantic importance similar to the fixation count. It generally indicates that a specific AOI is semantically informative to the viewer (Holmqvist et al., 2011). In other words, the longer the total time of watching is, the more informative it is to the viewer. These measures are defined with details in the results section.

In sum, this study attempts to investigate the effects of using visuals on test takers' performance on listening comprehension tests. The researcher compares scores of test takers' performance in three types of listening texts (video, still photos, and audio only) to find out whether their scores are affected by any of these texts. Then test takers' own perceptions of the various text types are extracted using a cued retrospective report in which they can view their own eye-movement behaviour as cues for the verbal recall. These reports are also used to track any evidence of cognitive processes used by the test takers during processing the listening texts and answering the test questions. Finally, an examination of the test takers' eye-movements is conducted in an attempt to link their viewing behaviours to: A) their performance on the three text types, and B) their explicit perceptions of the different listening test contexts.

Research questions

The research questions that guided this study are:

Research question 1:

- To what extent does the performance of L2 test takers differ on listening texts with a] video texts, b] still photo texts, and c] audio-only texts?

Research question 2:

- How do L2 test takers perceive the visual information in the a] video texts and b] still photo texts when processing and answering the comprehension questions to theses texts as indicated by the cued retrospective report? ○ Is there any correlation between the test takers' perceptions and their test scores?

Research question 3:

- What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualised texts?

2.6. Summary of chapter two

This chapter presented a review of the literature related to the current research. In particular, the chapter was organized in a way to review from the general area of L2 ability to the more specific area of L2 listening ability and explored the different definitions of this ability in the literature, the models and processes of listening, and various classifications of types of listening. From the area of L2 listening ability the focus shifted to the area of assessing it, and explored the most important qualities of listening tests like authenticity and construct validity, then a review of academic listening was presented as it is the type of listening that is in focus in this study, followed by the factors that can affect the performance in listening test, including various characteristics related to the listener, the text, and the tasks. As the core of the present study is the use of visual materials in L2 listening tests, it was important to discuss the types of visuals used in L2 listening, presenting a classification of these types and review the previous research that was done in the field of L2 listening supported by visual materials. Finally, the chapter concludes by reviewing the methodological approaches of investigating L2 listening with visual materials, discussing two important tools, namely cued retrospective reports and eye tracking technology, and exploring the possibility of using these tools to investigate the cognitive processes that may underlie the process of listening comprehension. The next chapter presents the methodology used for designing and administering the listening test used in the current study and describes the types of data collected via the test.

Chapter Three Methodology

The impact of using visual materials on performance in L2 listening comprehension tests has not yet been fully explored, therefore its use in language tests still very limited. In particular, still photo texts are the only type of visual materials currently in use in some standardized listening tests (e. g., TOEFL) but the impact of using this visual type is not known and more research is still needed. In addition, research on the impact of visual materials on test takers' performance in L2 listening tests had traditionally focused on the product of the test, that is, test takers' scores, while ignoring the process that led to the score, like their underlying cognitive processes. This study attempts to design an exhaustive study that covers areas related to the product of listening (scores), as well as areas related to the process of listening (perceptions and viewing behaviour that may reveal some insights about the cognitive processes used by test takers).

This chapter provides a description of the methodology used in the study. It introduces first the research design of the study, and then describes the nature of the participants who took part as test takers, and the materials that were selected and how they were used to develop the final version of the listening test, including the audio and the two visual text types. In addition, this chapter also explains the process of data collection, handling, and analysing.

3.1. Research design of the study:

This study was designed to collect mixed types of data in order to answer the three research questions. Below, the type of data collected is stated under its related research question.

To answer research question 1: To what extent does the performance of L2 test takers differ on listening texts with a] video texts, b] still photo texts, and c] audio-only texts?

[1] Data of listening test performance which are the total test scores and scores from the three sub-test, which comprise the three types of texts used in the test (i.e. video, still photo, audio-only).

To answer research question 2: How do L2 test takers perceive the visual information in the a] video texts and b] still photo texts when processing and answering the comprehension questions to theses texts as indicated by the cued retrospective report?

2.1: Is there any correlation between the test takers' perceptions and their test scores?

[2] Data of the participants' verbalization of their perceptions and indications if their inner cognitive processes during the test. These are collected via the cued retrospective reports, which use eye tracking data as the 'cues' for the cued retrospective reports in the form of heat maps, to help participants verbalize their perceptions. These data are then correlated to test takers' scores.

To answer research question 3: What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualised texts?

[3] Data from eye tracking, using three eye tracking measures (fixation count, dwell rate, and total dwell time) are used in order to independently analyse the records of the eye movement of each participant during watching the visualized texts (video and still photos). These data are then correlated to test takers' performance across the visualised texts [1], and also correlated to their perceptions and views of the visualised texts as elicited by the cued retrospective report [2]. Table 3.1 below summarizes the collected types of data.

DATA SET	DATA CATEGORY	DATA COLLECTION
TEST SCORES	Quantitative	Scores from the V, SP, and A sub-test.
VERBAL DATA	Qualitative	Cued retrospective report.

EYE TRACKING DATA	Quantitative	Eye tracking.

Table 3.1: Types of data collected in the study.

Overall, a mixed method research design was followed, which is methodological triangulation as proposed by Creswell and Plano Clark (2011). This includes the collection of quantitative and qualitative data. The quantitative data are gathered by the test scores and eye-tracking data. The qualitative data are collected via the cued retrospective report data, which immediately follows the performance of the test. Figure 3.1 below illustrates the triangulation design of the study.

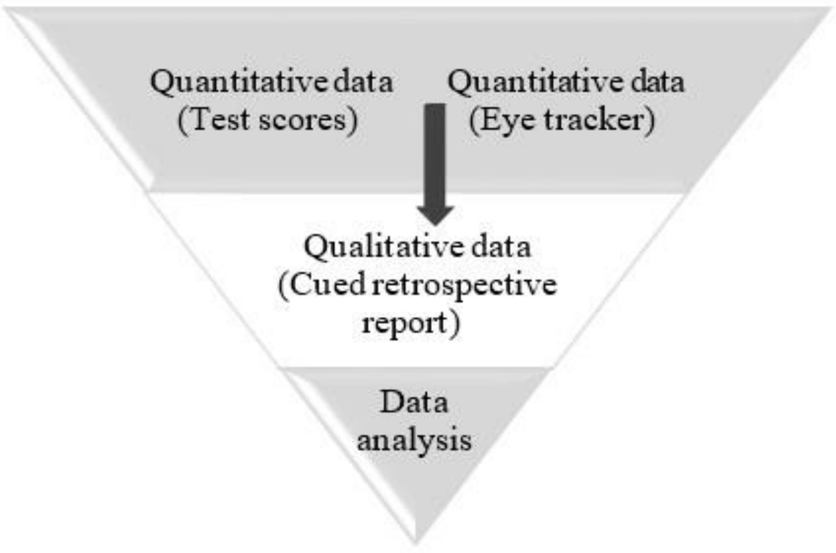

Quantitative data (Test scores)
Quantitative data (Eye tracker)
Qualitative data (Cued retrospective report)
Data analysis

Figure 3.1: Triangulation design of data collection.

On the bases of these data, the variables are identified to answer the research questions. Hence, the independant variables of the this study are the three types of listening texts (videos [V], still photos [SP], and audio-only [A]). Consequently, the dependant variables are consisting of three sets: (1) listening test scores (these are the scores of three listening subtests, V, SP, and A). (2) test takers‘ perceptions regarding the test (via cued retrospective report), and whether the visuals contribute to their comprehension during watching the text and answering the test questions. (3) test takers‘ record of eye-movemet (via eye tracker) during the visualised texts [V] and [SP].

3.2. Participants

The participants in this study were EFL students from two proficiency levels- intermediate and advanced. The intermediate group were international students enrolled in the pre-sessional course at the (CELT), that is the Centre of English Language Teaching at the University of York. These pre-sessional courses are aimed for undergraduate and postgraduate students who scored less than 6.0 in IELTS. The advanced group are mainly consisting of international postgraduate students enrolled in master and PhD programs at different departments in the University of York. Those are considered advanced because they had to have a score above 6.5 in IELTS in order to be enrolled in those programs. The rational for including different levels of non-native speakers (NNS) of English and from different cultural backgrounds is to represent the target population of proficiency tests (of which some will pass and some will not) who aim to study at universities were English is the main mean of instruction and communication.

The majority of participants are from different parts of Asia (i.e., Chinese, Japanese, Thai and Indians), the rest are a mix of Arabs, South Americans, and Europeans. Recruiting is done by the researcher who made short visits to the targeted participants in the CELT, who are the less advanced group. The more advanced are approached by sending e-mail requests to current students in different master and PhD programs in order to introduce the study, its purpose, and ask them to participate in it. Students who agreed to take part in the study were asked to sign a consent form (Appendix 2). Table 3.2 below provides background information about the participants. This information was collected by a short pre-test biographical questionnaire (Appendix 3).

YEL*	NATIVE	PROFICIEN	GENDER	AGE	MAJOR
	TONGUE				CY LEVEL

F (N =18) M (N = 12)	18 (n = 1) 19 (n = 1) 21 (n = 1) 22 (n =4) 23 (n = 9) 24 (n = 3) 25 (n = 3) 27 (n = 2) 30 (n = 2) 31 (n = 2) 36 (n = 1) 39 (n = 1)	Mandarin (n =12) Cantonese (n = 2) Arabic (n = 8) Malay (n = 1) Spanish (n = 3) Polish (n = 2) Russian (n = 1) Japanese (n = 1)	MA TESOL (n = 11) PhD Education (n = 3) MA Education (n = 5) Economics UG** (n = 3) Computer Science UG (n = 4) Health UG (n = 2) Politics UG (n = 1) Human rights UG (n = 1)	2 (n = 4) 3 (n = 6) 4 (n = 1) 6 (n = 1) 9 (n = 2) 10 (n = 5) 11 (n = 1) 12 (n = 1) 13 (n = 1) 14 (n = 1) 15 (n = 1) 16 (n = 2) 18 (n = 1) 19 (n = 1) 20 (n = 2)	Advanced (n = 19) Intermediate (n = 11)

Table 3.2: Test takers' biographical information (n = 30).

**YEL = year of English learning. ** UG = undergraduate.*

3.3. Materials

The listening text transcripts are taken and adapted from previous IELTS listening tests. IELTS is the International English Language Testing System, and it has very high standards for assessing the four language skills: reading, listening, writing, and speaking. The reliability of the listening test in IELTS, using Cronbach’s alpha is 0.90 for multiple versions used in year 2014, as reported on IELTS official homepage. Similar results were obtained in previous years. This figure (0.90) is regarded acceptable as a measurement of the consistency and reliability of a test (Hughes, 2003). The content validity of IELTS is also regarded high by many researchers (Bachman, 1995; Weir, 1990), however, some researchers (Field, 2009; Ockey, 2007) suggest that the format and construct of the listening test may need further revision, as research results indicate that test takers tend to perform better on contexts similar to real-life. Here, the *listen-read-write* format followed in almost all high-stake tests does not provide this environment. However, as the listening texts used in this study were originally designed for an audio only listening test, one might raise a concern that the role of visuals can be limited since comprehension of the text doesn't rely on it. Hence, it is important here to stress the fact that the main purpose of this study is to investigate the effect that visual materials might have on the behaviour of test takers, even if it is not a prerequisite for comprehending the text and answering the test questions. The main point is to explore whether the visual aspect would improve or hinder test takers' performance.

Nevertheless, in the current study the IELTS texts are used since their overall reliability and validity levels are high. The chosen listening texts were first recorded visually using a digital video camera, and designed specifically for this study. These texts were manipulated and reproduced to reflect the listening construct employed in this research without changing their basic content, by filming them in appropriate conditions that represent real lectures' environment like auditoriums or classrooms fitted with data-show screens in order to be able to display the Power Point slides for each text. Two different visual formats were produced beside the audio only format:

First, a video format, showing a lecturer delivering his/her speech to students, using normal body language and normal academic presentation skills, in which the language used is academic, though it may include some stuttering or miscues which may happen naturally by any speaker. The video also serves to capture the surrounding environment where the lecture took place without showing the students area.

Second, a series of still photos of the lecture are displayed at regular intervals. The timing of changing each photo in the still photo texts differed from one text to another to be in balance with the type of audio information being played. These still photos were basically created by taking screen-shots from the videos, and they served mainly to provide information about the physical setting of the lecture and to show the speaker. The next section describes in more details the process of developing and producing the L2 listening test used in the study, explaining all the steps taken from the beginning of that process until the production of the final version of the test.

3.4. Developing the listening test

In order to develop the visual materials needed for the test, the researcher approached a number of native English language speakers who work at the University of York, by sending them request e-mails asking them to participate as lecturers. Six teaching staff (of which 2 males and 4 females) at the university of York had agreed to participate in recording the videos. Initially, nine videos were produce using Sony full HD video camera, recorded by a specialist cameraperson. In each video, the main scene views the upper halve of the lecturer besides a screen shows some content information like an outline of the lecture or a diagram. After recording the nine texts in video format, the other two formats (still photos and audio-only) were produced. The still photo format was created by playing the video for each text and taking several screen shots along the video, and then fitting these photos with the sound version of the text. This was done by Windows Movie Maker application. It was decided to take screen shots from the video version instead of taking pictures of the speakers during the recording sessions in order to have exactly the same frames of the lecture scene in both video and still photo texts. The angle of taking pictures specifically for still photo texts might differ from the one used in the videos, which might affect the viewing patterns of test takers, and consequently affect the accuracy of data collected via the eye tracker. The audio-only format was created by extracting the sound of each video text using VLC Media Player application, which allows for converting video files to a high quality audio-only format. Hence all three text types are based on the same original video recording and share exactly the same spoken input. The test was then assembled and produced via Experiment Builder, a graphical programming environment that is used to produce different types of experiments using complex visual and auditory stimuli with very high levels of accuracy.

Piloting the listening test

The initial version of the test consisted of nine listening texts (3 video, 3 still photos, and 3 audio-only), and 45 questions (5 questions for each text). In the original version of the IELTS texts, each text has 10 questions. For practicality reasons, only 5 questions were chosen for this study. The process of choosing five questions out of ten for each text was done by the researcher on the bases that these five questions should test the test takers' comprehension of information distributed along the text and assess understanding of both general ideas and details. These choices were then revised by an experienced listening test designer and slight changes were adopted accordingly. Two item types were chosen: multiple choice questions and gap filling. More details on the reasoning of selecting these types below (see p. 99).

Moreover, other procedural issues related to the design of the test were also considered at this stage. Firstly, the number of times each text should be played. Is one time enough for test takers? Or, should the text be played two times to give test takers a chance to chick for their answers? Unfortunately, this issue is virtually neglected in the literature, as there is no sound argument in favour of one procedure over the other. Therefore, it is up to test developers to decide which procedure they will follow. In the current study, mainly for practicality reasons, every text was played only once in order to fit all the listening texts in a reasonable amount of time. Besides, it is argued that this procedure reflects the characteristics of target academic environment, where students can normally listen to real lecture for one time (Field, 2013). Double play of listening texts is normally used in listening comprehension tests in order to compensate for the lack of supporting visual materials, which normally accompany academic lectures (Geranpayeh and Taylor, 2008), hence, the decision of playing the text only once in the current study was supported by the fact that the visual aspect is available in two out of three text types (video and still photo texts) in the listening test.

Secondly, the issue of question preview was addressed. Should the questions be presented before the text is played, or should they be given after listening to the text? Once more, no consensus is available in the literature. This can be attributed to the different nature and purposes of different listening tests. Sherman (1997) explained this issue very clearly saying “question preview may affect comprehension positively by focusing the attention or supplying information about the test, or negatively by interfering with the subjective comprehension processes, increasing the burden on the attention or imposing shallower processing” (p. 185). However, some researchers (Buck, 1991, 2001; Wagner, 2010) support viewing the test items before listening to the text. For instance, Buck (2001) found that question preview “does seem to have a positive psychological value for test takers, and conversely the uncertainty of not knowing why they are listening has a negative psychological effect” (p. 137). Field (2013) argues that preventing test takers from previewing questions for texts that last for more than one minute can be threatening to the cognitive validity of the test. Thus, for this study, because all the used texts last for at least 3 minutes, it was decided that test takers should have the chance of previewing the questions before each text is played. After listening to each text, the five questions are displayed in the screen, so the test-taker can have a chance to navigate and review them before submitting their answers and going to the next text.

Here are the characteristics of the listening test used in the pilot version in this study presented in Table 3.3, which is based on test specification format developed by Bachman and Palmer (1996).

COMPONENT	DESCRIPTION
PURPOSE	To measure the ability of L2 test-taker to understand visual and verbal information presented in academic lectures. The texts are either enhanced by visuals (still photos or videos), or presented as audio-only.
CONSTRUCT	The ability to process and comprehend information provided by acoustic and visual input from academic lectures in university settings where English is the main language of communication.
SETTING	Eye-tracker laboratory.
TIME	Two hours (inclusive of the interview time).
INSTRUCTIONS	This test consists of nine listening texts. In each text you will watch (listen to, for the audio-only texts) a short introductory university lecture. These nine texts will be presented in different formats, where three of them will be enhanced by videos, another three texts will be enhanced by still photos, and three with audioonly format. You will listen to each lecture one time only. You can have a quick view to the questions for each lecture before you watch it. After watching each part, you will be given five comprehension questions. Answer all questions. The test lasts for one hour, so be sure to use the time wisely.

INPUT AND EXPECTED RESPONSE	Nine listening passages are presented to test takers. These passages are supported by videos, still-photos, or audio-only format. The lectures are designed as an introductory level academic lectures, each presents a different topic that does not require specialized knowledge in the field. After watching each lecture, test takers are expected to answer five comprehension questions. These are different question types like multiple-choice questions (MCQ) with a number of short options, or gap filling questions (GF) like shortanswer questions with no more than three words to answer, or diagram labelling with a visual diagram, or sentence completion using information from the text.

Table 3.3 Continued

COMPONENT	DESCRIPTION
SCORING METHOD	All questions are scored automatically. A personal revision of the answers is done by the researcher in order to deal with spelling problems and assign the answer a score if it was correct with simple spelling mistakes. Every correct answer is assigned a value of 1. Incorrect answers are assigned a value of 0.

Table 3.3: Specifications of the listening test for the pilot study.

As mentioned earlier, all the texts were chosen from IELTS, and these texts were presented as introductory university lectures in different, general topics (i.e., History, Health, Architecture, Business, Psychology, sports, Zoology, Social Science, and Induction day lecture). All these lectures presented non-specialist topics, which did not require previous specialized topical knowledge or specific mastery of terminologies in order to understand them. Figure 3.2 below shows some screenshots from four lectures that were used in the piloting stage of the listening test.

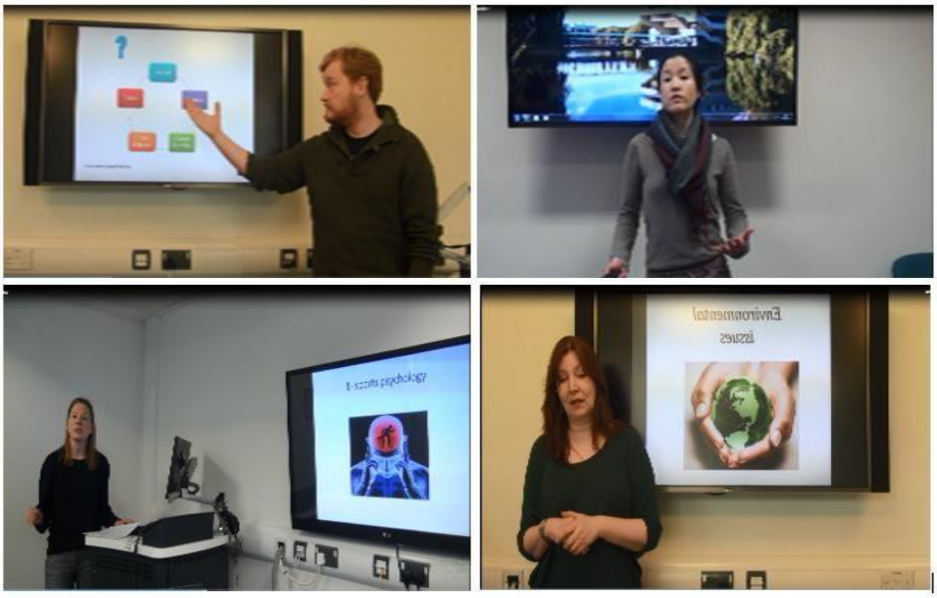


Figure 3.2: Screenshots from some of the first videos developed for the listening test.

Questions used to assess test takers' comprehension of the above texts were mainly of two types: gap filling (GF) and multiple-choice questions (MCQ). Gap filling is regarded in the literature as having a face validity value by resembling note-taking that usually takes place in real-life situations of listening to a lecture (Field, 2009, 2013). It also serves to test the understanding of clearly stated information (Buck, 2001) and that reflects what is stated in the construct used in this study (i.e., the ability to process and comprehend information provided by acoustic and visual input). The answers to gap-filling questions should not be more than three words in length in order to minimize the writing burden. A decision has been made regarding right responses with spelling mistakes or different wording to be accepted as long as the test taker demonstrates full understanding of the question and information required. According to Field (2009), strict adherence to correct spelling and wording can result in disqualifying many test takers for the wrong reasons and raise a construct-irrelevant variable issue.

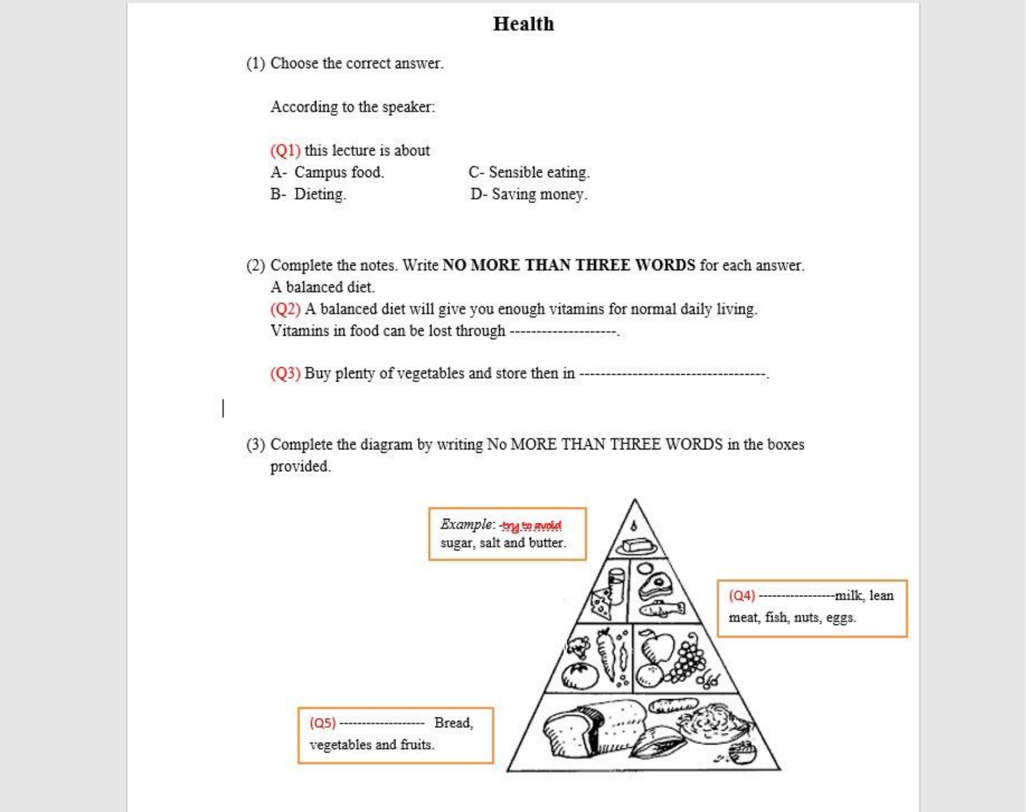
Multiple-choice questions were mainly used to elicit understanding of both main ideas presented in the lectures and details of the content of the texts. The use of MCQ also reflects the construct of the current study by testing general and detailed information, whether this information is provided verbally or visually. The MCQ used consist of a stem and vary in the number of choices, with some display three choices, and others display four short choices, besides only one instance of six choices with two correct answers to choose.

The test lay out in the pilot stage was as follows: A review of the five questions for each text was displayed for 20 seconds before the text started (that makes a total of 3 minutes for all the nine texts) so the participants had the chance to know the type of questions and type of information required to answer the test items. After that participants start listening to the text. (for the pilot, form A in Table 3.4 was used).

	V	**SP**	**A**	**V**	**SP**	**A**	**V**	**SP**	**A**
FORM A	Health	History	Induction day	Social Scienc	Business	Zoology	Psychology	sports	rchitectur Ā

Table 3.4: Layout of the texts used in the pilot.

After playing each text, the screen displayed the questions again, and at that time the test takers were able to enter their answers. All questions related to one text were displayed in one page that did not require any scrolling down. Test takers had to tap Enter whenever they wanted to move to the next question, or tap Back whenever they wanted to modify a pervious answer in the same page (Figure 3.3). Once they finished answering the test items they were able to move to the next text even before the end of time allowed for answering the questions. Maximum time for answering the test questions for each text was initially 90 seconds in the pilot. The total amount of time for playing the nine texts was 40:44 minutes. Adding to that 20 seconds for reviewing questions before each text (i.e., 3 minutes), and 90 seconds maximally for answers (i.e., 13:30 minutes) makes the total time of each test session 57:14.



Health

(1) Choose the correct answer.

According to the speaker:

(Q1) this lecture is about
A- Campus food. C- Sensible eating.
B- Dieting. D- Saving money.

(2) Complete the notes. Write **NO MORE THAN THREE WORDS** for each answer.
A balanced diet.
(Q2) A balanced diet will give you enough vitamins for normal daily living.
Vitamins in food can be lost through --------------------.

(Q3) Buy plenty of vegetables and store then in -----------------------------------.

(3) Complete the diagram by writing No MORE THAN THREE WORDS in the boxes provided.

Example: -try to avoid sugar, salt and butter.

(Q4) -----------------milk, lean meat, fish, nuts, eggs.

(Q5) ------------------ Bread, vegetables and fruits.

Figure 3.3: A screenshot of the question page displayed for the test takers.

The main purpose of the pilot was to test the overall design of the listening test and to practice the use of the eye tracker with real participants. Participants in the pilot were five international PhD students from different departments. Comments and recommendations from those participants were recorded along the trials, which unexpectedly lasted around three hours per each participant, because of their detailed feedback on each stage of the test which proved to be valuable for shaping the main version of the test. Findings from the pilot revealed no significant difference in the performance of test takers in the three text types. However, the limited number of participants and the similarity of their proficiency levels have probably caused this result. On the other hand, their perceptions provided basic concepts to the researcher which were related to both video and still photo texts, as seen to be containing helpful and distracting aspects. These perceptions helped in shaping and categorizing the verbal data later in the main study.

Main version of the test

A revised version of the test, after piloting with the five volunteer test takers, was adopted. Based on the reviews collected from those volunteers, a decision was made to shorten the test up to six texts only (2 video, 2 still photos, and 2 audio-only) for practicality reasons, as during the pilot students expressed signs of fatigue towards the end of the test because of spending long time their heads positioned in the desktop mount (see Figure 3.4). Three texts were eliminated (History, Induction day, and Psychology) because, according to the volunteer test takers in the pilot, those texts were the hardest or were longer than the rest of the texts (e. g., History text lasted for 06:08 minutes). The end result was six listening texts as shown in Table 3.5. Also, the question answering time was increased from 90 seconds to two minutes, as during the pilot test takers stated that the time was not enough for them to review and think about the answers for all the five questions. In addition, the six remaining visual texts were re-filmed (Architecture, Business, Health, Social Science, Sports, and Zoology), in order to correct the mistakes happened in the first recording, i.e., the speakers in different videos appeared standing in different positions from the slides in the background (Figure 3.2), which might have an effect on the viewing patterns of test takers, making them vary largely. As a result, the data collected from the eye tracker cannot be regarded reliable across different videos. Therefore, the positions of speakers were unified across all six texts, and a second video recording took place with all the speakers appear standing in the right side of the screen, and the slides appear in the left side of the screen.

Text	Architectur	Business	Health	Social Scienc	Sports	Zoology
Time	03:40	03:59	03:51	05:20	05:12	04:40
Q preview time	00:20	00:20	00:20	00:20	00:20	00:20
Q answer time	02:00	02:00	02:00	02:00	02:00	02:00

Total test time	41:42 minutes

Table 3.5: The length of each text in minutes as used in the main study and the total amount of time of taking the test.

As explained earlier, two types of questions were used with the six listening texts. The distribution of the multiple-choice questions and the gap filling questions across the listening texts is presented in Table 3.6. Texts used in the main study and their questions are listed in Appendix (1).

Text	Gap-filling	MCQ
Business	Q1, Q2, Q3, Q4, Q5.	
Architecture	Q3, Q4, Q5	Q1, Q2
Health	Q2, Q3, Q4, Q5	Q1
Social Science	Q1, Q2	Q3, Q4, Q5
Sports	Q1, Q2, Q3, Q4, Q5	
Zoology	Q1, Q2, Q3, Q4, Q5	

Table 3.6: The distribution of MCQ and GF items in the six listening texts.

Main test administration

The main test was administered in a within-subject experimental design. The choice of this design was based on the fact that it avoids the possible individual differences in test takers' performance. Each participant took the test individually. The process of testing participants (*n* = 30) lasted along two months. The researcher started collecting data by administering the test to each participant replied to the recruiting email by selecting a suitable time from the Doodle poll provided.

Once a participant arrived at the eye tracker laboratory, the researcher started explaining the process of taking the test and reminded them that the test is followed by an interview. After that participants were asked to sign a consent form (Appendix 2). The eye tracking laboratory basically consists of a small room, equipped with EyeLink 1000 eye tracker device (average accuracy 0.25°– 0.5°, resolution 0.01° RMS) supported by Experiment Builder, which is a software package used to design different types of experiments for research purposes. The equipment consists of two computers, the first is called the Host PC, on which the record of eye tracking is performed at 1000 sample rate per second, and stored in a data file. The Host PC is connected to a display screen, on which the actual experiment is displayed to the test takers. An infra-red light source and a video camera are fitted beneath the display screen and used to capture the viewers' pupils and record their watching patterns. There is also a desktop mount, that is used to rest the viewers' head in front of the display screen and minimise any head movements or shaking during the viewing time which help to record the watching patterns with high accuracy. Figure 3.4 shows the eye tracker lab.

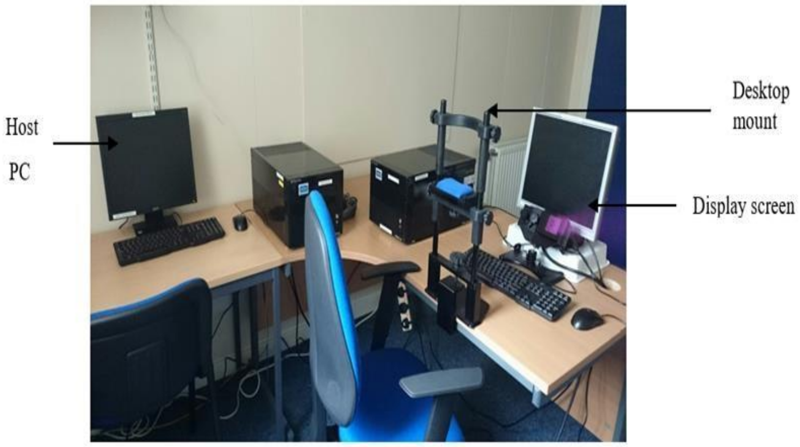


Figure 3.4: The eye tracker lab.

The sessions in the main study were organised in the following way:

First, the procedure of the test was administered to the participants. The researcher explained in some details what the test was about, how many texts they were supposed to watch, and the number of questions and how to navigate between them, without giving special attention to the importance of using the eye tracker in order to make sure that the test takers behave normally during watching the visual texts. A short biographical pre-test questionnaire was given to each participant (Appendix 3) in order to collect some background information (provided in Table 3.2 above). Before staring the test, the test takers were given the opportunity to watch two warming up visual texts, one in video condition and another in still photo condition with a short preview of the questions and then displaying them again at the end of the texts, in order to prepare them to the process of test taking. Instructions for taking the test were also provided on the screen at the very beginning of the test.

Also during this initial phase of the test, eye calibration was performed in order to identify the eye movement characteristics of each participant and calibrate them. Here, the participant is seated about 60 cm away from the display screen, with his/her head on the desktop mount which is supported with the head and chin rest. After making sure that the participant is setting comfortably, a nine-point calibration process starts in order to ensure the accuracy of each participants' eye measurements. The nine points are distributed to cover all parts of the screen as shown in Figure 3.5, in order to make sure that the participants do not have problems with viewing any part of the screen.

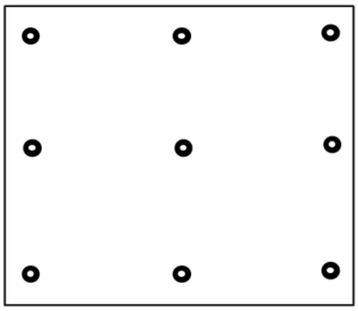


Figure 3.5: Distribution of nine-point calibration.

If the calibration was completed successfully, the participant can proceed to the next level, which is test taking. If the calibration was not successful, the participant is excluded. For example, four participants in the main study were disallowed from taking the test because of a calibration problem. All nine calibration points were fixated with an error greater than 1°. This indicates that there is a large difference between the actual position of the displayed points and the position where the participants fixated their eyes. In this case the calibration cannot be accepted because the error is too high for useful eye tracking, and consequently the results will not be accurate. Figure 3.6 shows examples of good and poor calibrations.

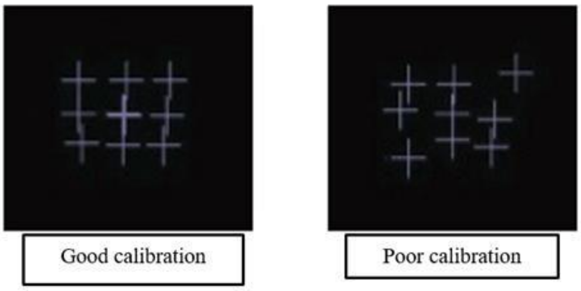


Figure 3.6: Examples of good and poor calibrations.

Sampling rate of the eye tracker determines the frequency with which a data set is recorded and specified in Hertz (Hz) but are easily converted into time measurements by dividing the value into 1000 (milliseconds). (e. g., a sampling rate of 1000 Hz records a data point every 1 millisecond: 1000/ 1000 = 1). In essence, there is a trade-off between the sampling rate and the degree of movement that the participants can engage in. The higher the sampling rate, the more stable the participants' head must be.

Every individual participant took a listening test which consisted of the selected six texts, distributed on the three types of listening conditions (video texts, still photo texts, and audio-only texts). In order to avoid the order effect that can be associated with presenting the texts in the same order to all test takers; the test was administered in three forms (A, B, C). Each form presented the same six texts but with a different order and a different choice of visual or audio-only input. Table 3.7 shows the distribution of listening texts and the visual choices in the three forms.

Form A	**V** Health	**SP** Social	**A** Business	**V** Zoology	**SP** Sports	**A** Architecture

Science

Form B	**SP** Architecture	**A** Sports	**V** Social Science	**SP** Zoology	**A** Health	**V** Business
Form C	**A** Zoology	**V** Sports	**SP** Business	**A** Social Science	**V** Architecture	**SP** Health

Table 3.7: Structure of the three forms of listening test. V = Video. SP = Still photos. A = Audio.

Second, taking the test. During this phase participants listen to the six texts as explained previously. They are given the free choice of looking at the screen or away. The participants are also given some papers and a pen to help them in the process of taking notes if they wish to. After answering the questions of each text, a drift check is performed in order to ensure that the participants' calibration is still accurate. The drift check is basically a one-point calibration check. It also provides the advantage of taking a short break between the texts if the test takers felt they needed to rest.

Third, establishing a basic cued retrospective report with the visualised texts. This is done immediately after completing the test. The completed data file stored in the Host PC is sent through an Ethernet connection to the display screen. This data is saved into an EDV file (EyeLink Data Viewer), which is a tool that allows for displaying heat maps. Heat maps are generated in the Data Viewer order to show test takers a record of their eye movement on the two visualised texts (Figure 3.7). These maps help to give a summary of the areas where test takers had focused their visual attention by different colour highlights, with the red colour reflecting a hot area, which means that high amounts of fixation had been placed on that zone. On the other hand, green colour reflecting a cold area, which mean low amount of fixation had been placed on that zone. These heat maps work as a cue that help test takers to explain their viewing behaviour, and give detailed reasoning for the way they watched each visual. For the main test, a heat map display for every visual text was performed (2 video texts, and 2 still photo texts) with every individual test taker. They were asked general guiding questions to facilitate the process of their verbalizations regarding the whole course of test taking. These cued retrospective reports were recorded using Parrot voice recorder software, which is a high-quality HD program that allows for recording, storing, and playing voice records.

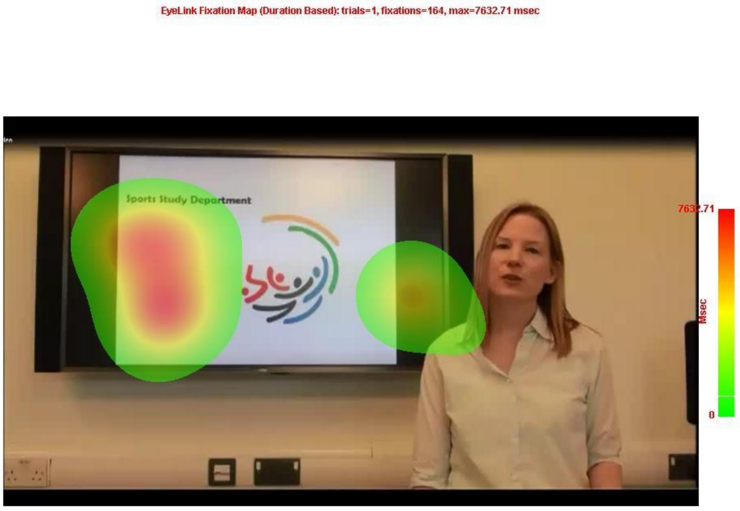

EyeLink Fixation Map (Duration Based): trials=1, fixations=164, max=7632.71 msec
Sports Study Department
7632.71
Msec
0

Figure 3.7: A screenshot shows the eye positions of one of the participants, as used in the cued retrospective report.

3.5. Types of data analysis

As specified earlier, quantitative and qualitative data are collected and analysed in order to answer research questions. The types of data analysis employed to answer each question are as follows:

Research question (1): To what extent does the performance of L2 test takers differ on listening texts with a] video texts, b] still photo texts, and c] audio-only texts?

To answer the first research question a repeated measure ANOVA was conducted in order to see if there was any difference in test takers' performance (*n* = 30) in the three listening texts enhanced by video, still photo, or audio-only texts.

Research question (2): How do L2 test takers utilize and perceive the verbal and non-verbal information in the a] video texts, and b] still photo texts when processing and answering the comprehension questions in these texts, as indicated by the cued retrospective report? Is there any correlation between the test takers' perceptions and their scores?

Verbal data collected from the cued retrospective report are transcribed and analysed to find out whether test takers used visualised data in the two visual conditions similarly or in a different way. Nvivo 11, which is a qualitative data analysis software was used for the analysis of this type of qualitative data. The procedure of transcription is explained in more details in Table 3.8.

Test takers	Duration of the verbal report	Word count of the transcribed data
Test taker 1	00:52:03	3,289
Test taker 2	00:47:33	2,985
Test taker 3	01:02:44	4,109
Test taker 4	00:58:18	3,537
Test taker 5	00:17:29	1,209
Test taker 6	00:56:49	2,849
Test taker 7	00:44:12	2,252
Test taker 8	00:37:40	2,408
Test taker 9	00:48:19	2,836
Test taker 10	00:28:50	2,147
Test taker 11	00:25:07	1,409
Test taker 12	00:47:48	2,783
Test taker 13	00:54:55	3,121
Test taker 14	00:33:42	2,038
Test taker 15	00:53:06	3,362
Test taker 16	01:04:38	4,160
Test taker 17	00:26: 12	1,946
Test taker 18	00:22:01	862
Test taker 19	00:39:54	1,047
Test taker 20	00:51:15	3,186
Test taker 21		

Test taker 22	00:40: 07	2,110
	00:48:11	2,690
Test taker 23	00:23:55	1,833
Test taker 24	00:56:48	3,729
Test taker 25	00:59:03	4,053
Test taker 26	00:33:45	1,181
Test taker 27 Test taker 28	00:46:46	2,051
	00:26:13	1,912
Test taker 29	00:29:38	2,066
Test taker 30	00:35:21	2,403

Table 3.8: Summary of collected verbal reports per test taker.

After the transcription process, the data were entered into Nvivo 11 to start the coding procedure. Table 3.9 explains the initial coding categories used to analyse the collected verbal data.

Category	Description of the category	Guiding question
Text type Video focus aspects	The preferred type of text: video, still photo, or audio only.	Which of the three types of texts you find more helpful?
	Aspects in video that participants focused on.	While watching the video, what aspects do you focus on?
Video focus reasons	Reasons for focusing on the stated aspects of videos.	Why do you focus on these aspects of the video?
Still photo focus aspects	Aspects in still photos that participants focus on.	While watching still photos, what aspects do you focus on?

Still photo focus reasons	Reasons for focusing on the stated aspects of still photo.	Why do you focus on these aspects of the still photos?
Video help aspects	Aspects of video participants find helpful (face, lip movement, gestures, etc.)	Do you find the video helpful? / What aspects you find most helpful?
Video help reasons	Reasons of finding the video helpful.	Why do you find these aspects helpful?
Still photo help aspects Still photo help reasons	Aspects of still photos participants find helpful.	Do you find the still photos helpful?
	Reasons of finding still photos helpful.	Why do you find these aspects helpful?
Video distract aspects	Aspects of video participants find distracting.	Do you find the video distracting?
Video distract reasons	Reasons of finding the video distracting.	Why do you find these aspects distracting?
Still photo distracting aspects	Aspects of finding still photos distracting.	Do you find still photos distracting?
Still photo distracting reasons	Reasons of finding still photos distracting.	Why do you find these factors distracting?

Table 3.9: Coding categories for the verbal data collected from test takers.

These coding categories were basically based on the perceptions and feedback collected during the process of piloting the study, and then filtered and revised during the main data qualitative analysis.

In order to answer the second part of research question two (Is there any correlation between the test takers' perceptions and their scores?), Spearman rank-order correlation test is conducted to find out if there is any relationship between the test takers' scores and their perceptions.

Research question (3): What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualised texts?

Three basic eye tracking measures are examined. These measures are fixation count, dwell rate, and total dwell time. The definitions of these eye tracker measures are presented in Table 3.10 according to Holmqvist et al., (2011).

Measure	definition
Fixation count	Is the number of fixations in the area of interest.
Dwell rate	Is the number of entries into a specific AOI per minute m^{-1}.
Total dwell time	Is the sum of all dwell times in one AOI over a trial.

Table 3. 10: Definitions of eye tracking measures used in the study.

Several statistical measures are used in order to answer research question 3. First, to answer the first part of the question (What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker) Paired-samples *t* tests were used to compare test takers viewing patterns in the video texts with their viewings in the still photo texts. In order to answer the second part of the question (Is there a connection between these patterns and test takers' performance in the test), each of the three eye taking measures was correlated with the video texts' scores and with the still photo texts' scores using Pearson correlation coefficient (*r*). Similarly, to answer the third part of research question three (Is there a connection between these patterns and their perception of the visualised texts), again Pearson correlation coefficient (*r*) was conducted to find out the type of relationship between the test takers' viewing patterns and their perceptions.

3.6. Summary of chapter three

This chapter presented the methodology used in this study in order to answer the three research questions. A mixed method design was adopted to tackle the different requirements of the research questions. More specifically, the study used data triangulation style as proposed by Creswell and Plano Clark (2011), that allows for simultaneous collection of quantitative data (listening test scores and eye tracking data) followed immediately with qualitative data (verbal data collected via the cued retrospective report). Besides, the chapter also presented a detailed explanation of the participants, materials used for the test, how the test was piloted, and how the main version of the test was administered. To finish, the chapter concluded with a short explanation of the types of data analysis for each research question.

.

Chapter Four Results

The purpose of this chapter is to provide answers to the research questions listed previously in the methodology chapter. The chapter is organised to present the results of data analysis of each research question, followed by a discussion section that covers the related topic. The types of data analysis are three sets, first, quantitative analysis of participants' test scores, which compares test takers' scores on three conditions of listening texts (video, still photo, and audio) in an attempt to answer research question one. Second, qualitative analysis of the cued retrospective report data that test takers' provided regarding the video and still photo texts, and that provides answer to research question two. Lastly, quantitative analysis of the eye tracking data which reveals the viewing patterns of test takers in the video and still photo texts, and that gives answer to research question three.

4.1. Research question 1:
To what extent does the performance of L2 test takers differ on a listening test with a] video texts, b] still photo texts, and c] audio-only texts?

The analysis of the first research question examined the extent to which the performance of test takers (n = 30) differed on the three listening test conditions (video, still photos, audio). More precisely, whether there was a statistically significant difference between their scores in the three text types. The maximum score for each participant was 30 (10 scores for each test condition).

To measure the difference in test takers' performance in these three test conditions (IV), one way within-subjects (repeated measures) ANOVA was conducted. First, descriptive analysis of the scores in the three text types presented in the listening test is reported in Table 4.1, in addition to a histogram that shows the distribution of test takers' scores in video, still photo, and audio texts (Figure 4.1). finally, the percentages of scores for each individual text in the three test conditions are presented in Table 4.2.

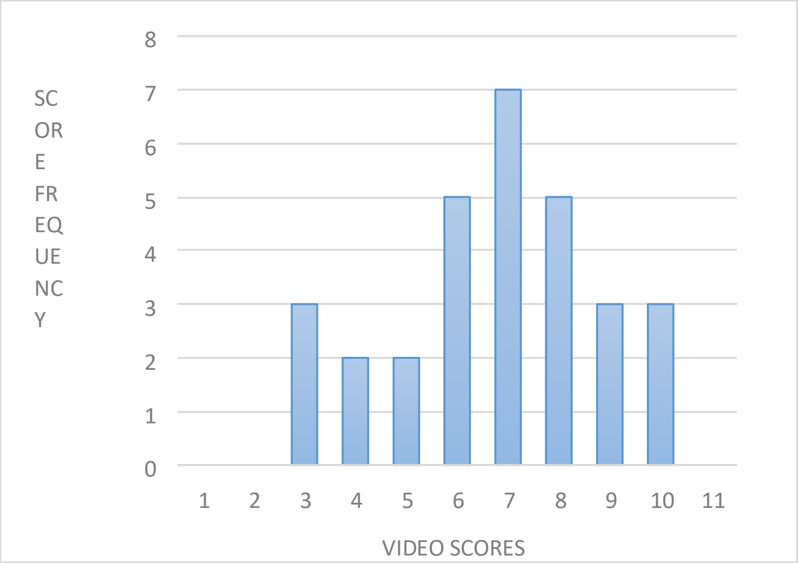
Text type	M	SD	Max	Min	Range	Skewness	Kurtosis
Video	5.77	2.17	10	2	8	-.309	-.537
Still photos	5.50	2.37	10	1	9	-.252	-.882
Audio	4.57	2.40	9	0	9	.153	-.753

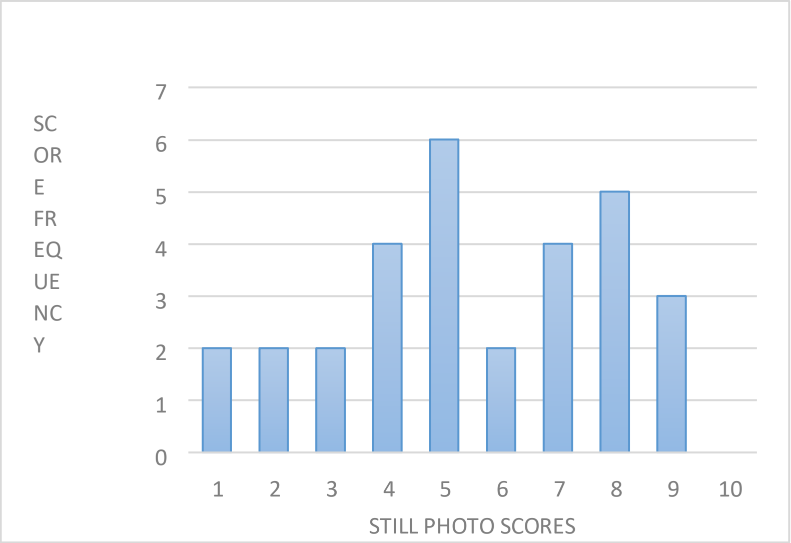
Table 4.1: Descriptive statistics of the scores in three text types.

As shown in Table 4.1, the highest mean was recorded for the video condition scores (*M* = 5.77, *SD* = 2.17), which means that 57.7% of the answers in this condition were correct.

A similar, though slightly lower mean was found in the still photo texts (*M* = 5.50, *SD* = 2.37), that is 55% correct answers. With the audio texts the mean was lower than the other two texts (*M* = 4.57, *SD* = 2.40), that is less than half the answers (45.7%) were correct.

Figure 4.1 presents visual display of the distribution of scores in the three text types. The histograms show evidence that the data was normally distributed as also indicated by the Skewness and Kurtosis values in Table 4.1. It is specified that values within ± 2 reasonably indicate normal distribution (George and Mallery, 2010).


SC
OR
E
FR
EQ
UE
NC
Y
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10
11
VIDEO SCORES


SC
OR
E
FR
EQ
UE
NC
Y
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10
STILL PHOTO SCORES

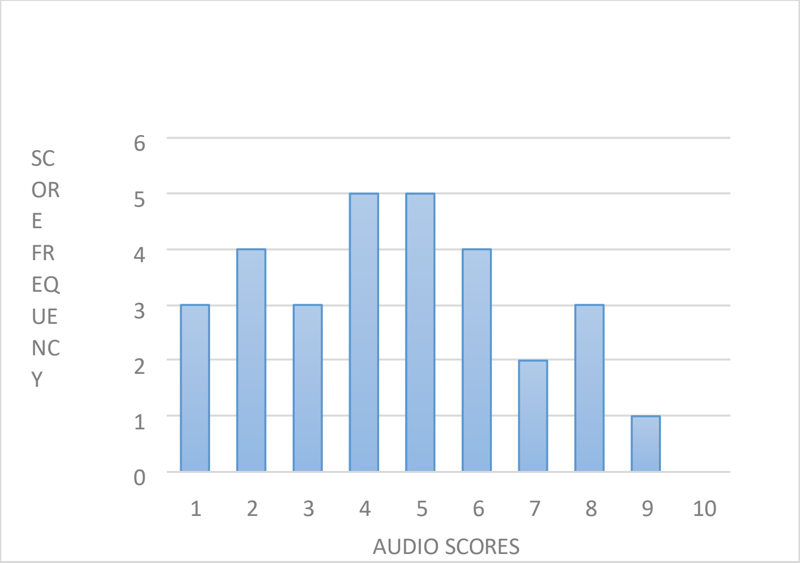

SC
OR
E
FR
EQ
UE
NC
Y
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8
9
10
AUDIO SCORES

Figure 4.1: Frequency histograms show the distributions of test takers' scores in the three text types.

Text title	Video texts	Still photo texts	Audio texts
Architecture	50%	38%	42%
Health	76%	64%	54%
Business	60%	54%	54%
Social Science	64%	56%	36%
Sports	66%	68%	56%
Zoology	40%	46%	32%

Table 4.2: Percentages of scores for each individual text in the three test conditions.

In Table 4.2, the percentages of each individual text show that in 4 out of the 6 texts (Architecture, Health, Business, and Social Science), test takers scored the highest in the video condition of each text, while in the other two texts (Sports and Zoology) the highest scores were in the still photo condition. Audio condition did not score higher than both the video and still photo with any of the texts.

In addition to the percentages of scores in each of the video, still photo and audio texts, the test takers' performance on the 30 individual items was also examined, in addition to the difference in their scores for each item.

Item number	*Item type*	*Video texts %*	*Still photo texts %*	*Audio texts %*	*V/SP difference %*	*SP/A difference %*	*V/A difference %*
Architecture 1	MCQ	49.8	29.8	45.3	20.0*	-15.5*	4.5
Architecture 2	MCQ	45.9	40.7	43.0	5.2	-2.3	2.9
Architecture 3	G-F	55.8	38.7	39.7	17.1*	-1.0	16.1*
Architecture 4	G-F	67.5	55.9	46.5	11.6*	9.4	21.0*
Architecture 5	G-F	33.2	22.9	34.3	10.3*	-11.4*	-1.1
Health 1	MCQ	54.9	44.9	40.6	10.0	4.3	14.3*
Health 2	G-F	78.9	67.9	55.3	11.0*	12.6*	23.6*
Health 3	G-F	66.5	50.2	46.1	16.3*	4.1	20.4*
Health 4	G-F	89.9	78.5	60.4	11.4*	18.1*	29.5*
Health 5	G-F	88.3	80.0	66.0	8.3	14.0*	22.3*
Business 1	G-F	45.3	42.2	44.8	3.1	-2.6	0.5
Business 2	G-F	60.2	55.4	50.7	4.8	4.7	9.5
Business 3	G-F	50.9	43.1	39.9	7.8	3.2	11.0*
Business 4	G-F	65.9	61.2	63.9	4.7	-2.7	2.0
Business 5	G-F	75.8	69.7	69.6	6.1	0.1	6.2
Social Science 1	G-F	72.6	66.1	62.9	6.5	3.2	9.7
Social Science 2	G-F	55.9	45.9	23.1	10	22.8*	32.8*
Social Science 3	MCQ	85.8	67.3	41.2	18.5*	26.1*	44.6*
Social Science 4	MCQ	38.2	37.0	27.9	1.2	9.1	10.3*
Social Science 5	MCQ	69.6	64.9	23.5	4.7	41.4*	46.1*
Sports 1	G-F	77.4	72.8	60.4	4.6	12.4*	17.0*
Sports 2	G-F	54.3	58.9	45.9	-4.6	13.0*	8.4
Sports 3	G-F	89.5	90.1	79.3	-0.6	10.8*	10.2*
Sports 4	G-F	53.7	55.6	44.9	-1.9	10.7*	8.8
Sports 5	G-F	58.0	67.2	51.2	-9.2	16.0*	6.8
Zoology 1	G-F	31.9	29.7	28.2	2.2	1.5	3.7
Zoology 2	G-F	29.8	49.9	37.9	-20.1*	12	-8.1
Zoology 3	G-F	47.0	52.1	35.5	-5.1	16.6*	11.5*
Zoology 4	G-F	45.5	51.5	28.6	-6.0	22.9*	16.9*
Zoology 5							

G-F	43.9	44.8	30.1	-0.9	14.7	13.8

*Table 4.3: Percentages of items scores in the three text types and the differences between them. * indicates at least 10% difference.*

The difference in test takers' performance in the 30 individual test items is examined by comparing the percentages of the scores in video condition to the scores in the still photo condition, then comparing the scores in the still photo condition to the scores in the audio condition, and lastly the scores in the video condition to the scores in the audio condition as presented in Table 4.3 above (V/SP, SP/A, V/A). In 8 out of the 30 test items, test takers in the video condition scored 10% or higher than the same items in still photo condition (Architecture 1, 3, 4, 5; Health 2, 3, 4; Social Science 3). The largest difference between video and still photo item scores was on item 1 of Architecture (20.0 %). This was a multiple-choice item asks about explicit information stated by the speaker but was not explicitly apparent on the slide. Only in one item the scores in the still photo condition were more than 10% higher than the video condition scores, and that was item 2 of Zoology, which was a gap-filling question asking about one of two types of food that whales consume and cause poisoning. This was also an explicit piece of information mentioned by the speaker. In the next comparison, test takers' scores in the still photo condition outperformed the scores in the audio condition in 13 out of the 30 items with a difference higher than 10%. These were items (Health 2, 4, 5; Social Science 2, 3, 5; Sports 1, 2, 3, 4, 5; and Zoology 3, 4). The highest difference scored in item 5 of Social Science with 41.4% difference. Only in the scores of two items (Architecture 1, 5) the difference was larger than 10% with the audio condition compared to the still photo condition (15.5%, 11.4% respectively). The first item was a multiple-choice question and the second was a gap-filling question, both asked about explicit information that was stated clearly by the speaker but was not described in the slides neither textually nor graphically. The last comparison, and the largest in its magnitude, was between the scores in the video and audio conditions. The difference in scores between these two conditions reached the highest percentage of score differences in item 5 of Social Science (46.1%). This item was a multiple-choice one, and it was asking a general question about how the Americans would choose to spend an extra day of the week. This information was not displayed in any shape in the slides and was only stated by the speaker. In total, the scores of 16 items in the video condition were more than 10% higher than the scores of the same items in the audio condition. Difference was found in items from all 6 texts, more specifically in the following items (Architecture 3, 4; Health 1, 2, 3, 4, 5; Business 3, Social Science 2, 3, 4, 5; Sports 1, 3; and Zoology 3, 4). None of the items in the audio condition scored higher than 10% from the video condition.

Difference in test takers' performance in video, still photo, and audio texts:

To analyse the results of test performance further, repeated measures ANOVA was conducted with an alpha level of .05, in order to inspect whether the degrees of difference among the test takers' scores in the three test conditions examined in Table 4.1 above are of any statistical significance. Table 4.4 shows the results of this analysis.

MAUCHLY'S TEST OF SPHERICITY			EPSILON		
APPROX. CHI-SQUARE	*df*	Sig	Greenhouse-Geisser	Hyunh-Feldt	Lowerbound
6.307	2	.043	.832	.877	.500

Table 4.4: Mauchly's test of Sphericity.

As repeated measures ANOVA was used, assumption of sphericity had to be made. It assumes that the relationship between the three test conditions is similar (Field, 2012). Mauchly's test was used in order to test this assumption and the results indicated that the assumption of sphericity had been violated with a statistically significant difference, χ^2 (2) = 6.30, *p* = .043. As the sphericity score in the Greenhouse-Geisser test was > .75 Therefore, degrees of freedom must be corrected using the Huynh-Feldt estimates of sphericity (ε = .87).

HYUNH-FELDT	*df*	*F*	Sig
	1.75	3.89	.032

Table 4.5: Test of within subjects effect (Hyunh-Feldt) to correct sphericity.

The results as appear in Table 4.5 show that there was a statistically significant difference in the test takers' performance in the three test types, *F* (1.75, 50.84) = 3.89, *p* = .03).

In order to know where exactly this difference exists, a Bonferroni post hoc test was conducted and the results revealed that the test takers' scores in the audio texts (*M* = 4.57) were significantly lower (*p* = .036) than their scores in the video texts (*M* = 5.77). There was no statistically significant difference between the still photo texts scores (*M* = 5.50) and the video texts scores (*p* = 1.000), nor between the scores of the still photo texts and the audio texts (*p* = .273). The results are presented in Table 4.6 below.

PAIRWISE COMPARISON		Sig
V A	S. P	1.000 .036
S.P A	V	1.000 .273
A	V S. P	.036 .273

Table 0.6: Results of repeated measures ANOVA comparing the scores in the three conditions, Video texts (V), Still photo texts (SP), and Audio texts (A).

As shown in Table 4.1 above, the mean scores of the video texts (5.77) was numerically higher than the mean scores of both still photo (5.50) and audio texts (4.57).

Comparing the highest mean with the lowest indicates that the test takers' scores in the video texts were 12% higher than their scores in the audio texts, and this difference is statistically significant as the results of the repeated measures ANOVA revealed. On the other hand, the scores of the video texts were 2.7% higher than their scores in the still photo texts, resulting in a non-significant difference between the two. In addition, the scores in the still photo texts were 9.3% higher than the scores in the audio texts, indicating a numerical difference only, and it was not statistically significant. In other words, the only statistically significant difference was found between the scores of the video texts and the scores of the audio texts.

In terms of the first research question "to what extent does the performance of L2 test takers differ on listening test with a] video texts, b] still photo texts, and c] audio-only texts?" the overall scores of the three text types presented evidence supporting the notion that the video texts did contribute to improving the test takers' scores in the L2 listening test. The performance of the test takers in the video texts was similar to their performance in the still photo texts, but higher (at a statistically significant level) than their performance in the audio texts. While the overall scores of the video texts were statistically higher than the audio texts scores, indicating that the visual component of the video texts may have contributed to improve the test takers' performance, it is not clear from the data which aspects of the visual components of the video texts might have led to improving the overall comprehension among test takers of the video texts and consequently contributed to better performance on the test items. Therefore, the different components of visuals that were present in the video texts (and some were also present in the still photo texts) but were not present in the audio texts are discussed in the following section in relation to their hypothesized effects on the performance of L2 test takers as presented in the literature.

First of all, it should be noted that the test items for all six texts were originally designed for an audio-only text and were not manipulated in any way to give advantage to the two visual texts. In other words, attention to the visual components was not a prerequisite to answer any of the test items. However, the visual information might have been helpful to answer some of the test questions in one way or another. For instance, for some test takers, the visual information might complement the audio information (Suvorov, 2013), or it might even provide redundant information which can be very important for confirming the test takers' processing of the audio input (Wagner, 2006b).

However, despite the results showing the general superiority of the video condition in terms of the scores, it should be noted that the helpfulness of video texts might not be the same across all these texts. In other words, it is conceivable that some test takers benefitted from a specific visual component in one of the video texts, but found the same component distracting in the other.

Several components of the visualised texts may have been beneficial for the test takers in processing the audio information and answering the test questions. One of these visual components is the tables and diagrams presented in the Power Point slides of some of the texts. The inclusion of tables in the video and still photo versions of some texts may have been supportive for the test takers in their comprehension of important parts of the audio input and can also help to answer some test questions. For instance, test takers scored 28% higher in the video version of Social Science text than in the audio version, and only 8% higher than the still photo version. In this text, there was a table that displayed the basic categorisation of the different social characters of the human beings described in the text. The same table is also displayed in the text questions and used to answer items 1 and 2. It is possible that the presence of the table while the speaker explicitly describes each item in it may have helped the test takers to answer the first two questions correctly even if they did not completely understand all the content of that table.

Another component of the visual texts that may have possibly contributed to achieving higher scores in the video and still photo texts is the display of photographs. In the Zoology text for instance, there was a picture of a group of dolphins stranded on a beach while the speaker was explaining the meaning of "mass strandings". Test takers scored in the still photo version of this text 14% higher than the audio version, and only 6% higher than the video version. The display of this photograph may have been useful to the test takers in comprehending, from the very beginning of the text, the general meaning of "mass stranding" which was a new term for most of them, and that might have enabled them to make accurate initial hypotheses about the content of the text. It can be said that the picture served as an activation to the listeners' background knowledge (Ockey, 2007) and provided a facilitative environment for better comprehension (Rubin, 1995). In the case of audio texts, test takers who do not know the meaning of stranding had probably wasted more time in attempting to understand what the speaker was talking about because there was no such visual clue to assist their comprehension.

As the previous two components (tables/ diagrams and photographs) are both present in the video as well as the still photo texts, the speakers' body language and their kinesic behaviour can only be observed in the video texts. This component of the video texts may have affected listeners' comprehension in a positive way. For example, in the video version of Architecture text test takers scored 12% higher than in the still photo version of the same text. It is important to note that this text did not display any tables or noteworthy photographs that are directly linked to the answer of any test item. There were several pictures that are generally related to the topic of architecture and environment but was not specified for the content of this lecture. The speaker, on the other hand, explained the process of building an environmentally friendly house and performed some gestures to explain several features of its design, like using her hand to draw an imaginary horizontally inclined line to describe the sloping nature of the land on which the housed was built. This can be a difficult idea to imagine if the test-taker does not know the meaning of a particular word like "slope", therefore, this was a potential instance in which gestures were useful in the delivery of the meaning and potentially made it easier for the listener to visualise the situation. However, the speakers' gestures may have been useful for the listeners in subtler ways as well. For instance, Von Raffler-Engel (1980) claims that gestures and other kinesic behaviours can reinforce the linguistic message, while Antes (1996) explains how gestures can serve different purposes like providing emphasis and clarifying ambiguous meanings, and Wagner (2006b) argues that gestures and body language can also serve to better mirror real life situations. In addition to these components that might have played a major role in rising the scores of the video and still photo texts, many researchers (Baltova, 1994; Dunkel, 1991; Suvorov, 2013; Wagner, 2002, 2006b) have described how the inclusion of visuals in listening texts can serve to boost the positive attitudes of the test takers and make the test tasks interesting.

Several explanations can be proposed to explain why test takers performed better in the video and still photo texts than in the audio texts. In order to investigate what aspects in the visualised texts may have been specifically helpful to the test takers in this study and what aspects they may have found distracting, the following section will provide the results of the cued retrospective report in which the test takers reported their use and their perceptions of the several components of the video and still photo texts.

Summary of research question one:

To address research question 1, repeated measures ANOVA was conducted to compare test takers’ scores in the three test conditions, Video, still photos, and audio texts. The results of this analysis revealed that video condition scores the highest mean in comparison with the other two conditions. There was a statistically significant difference $p < .05$ between the scores of test takers’ in the video condition and the scores in the audio condition. This result indicates that the video texts had some effect on the performance of test takers. No statistically significant difference was found between the scores of the still photo condition and the audio condition, nor between the scores of the video condition and the still photo condition. Several interpretations were discussed in an attempt to explain the reasons why test takers achieved the highest scores with the video texts.

The following section presents the results of the qualitative analysis of the cued retrospective report data in an attempt to spot the light on how the test takers perceived the visual texts (video and still-photo) in the L2 listening test.

4.2. Research question 2:

How do L2 test takers perceive the visual information in the a] video texts and b] still photo texts when processing and answering the comprehension questions to theses texts as indicated by the cued retrospective report?

2.1: Is there any correlation between the test takers’ perceptions and their test scores?

The overall purpose of research question 2 was to explore how test takers use the visual information when watching the video and still photo texts based on evidence from their cued retrospective report data using heat maps. The main body of this question is split into two areas of inquiry bout L2 test takers perceptions: first, their perceptions when processing the visual texts, and second, their perceptions when answering the test questions. The sub question then investigates the relationship between these perceptions and test takers’ scores in the visual texts, video and still photos.

L2 test takers’ perceptions of visual information when processing the visual texts.

To answer the first part of this question (How do L2 test takers perceive visual information while processing the visual texts), the researcher carried out a qualitative analysis to the verbal data collected from test takers via a cued retrospective report immediately after finishing the listening test. This data consisted of the responses that test takers gave regarding their visual behaviour while watching the video and the still photo texts. The interview put the spotlight on the aspects that the test takers focused on (as revealed by the heat maps) while watching the video and still photo texts, and the reasons why they focused on those aspects and whether they found them helpful or distracting. The researcher also asked the test takers how they answered each individual test item in both the video and still photo conditions, and whether the visual aspects in these conditions played a role in their answers to the test items.

To find out the answer to the first part of RQ2, the researcher recorded the interviews data and transcribed them, and then analysed them using a computer software package for analysing qualitative data (Nvivo 11).

The analysis of the qualitative data revealed that the test takers found some helpful and distracting aspects in both video texts and still photo texts to varying degrees. Each testtaker was asked to comment on each individual video text (total of 2) and each still photo text (total of 2) that he/she encountered during the listening test. In addition to these aspects, test takers were also asked to explain their reasons for focusing on specific zones as appeared in the heat maps screen while watching the video and still photo texts.

The results reported in the following section are organised according to the helpfulness of each text type followed by results of the reasons for finding aspects of that text type helpful. More precisely, the helpful aspects of the video texts are reported first (Table 4.7), followed by the reasons for finding aspects of the video text helpful (Table 4.8). Then the helpful aspects of the still photo texts are reported (Table 4.9) followed by the reasons for finding aspects of the still photo texts helpful (Table 4.10). The same process is repeated to report the distracting aspects and reasons for each text type.

Helpful aspects of the video texts:

Table 4.7 summarizes the major helpful categories and their related aspects of the visual information in the video texts as reported by the test takers. The number of comments column specifies the total number of all the comments made on each of the helpful aspects in the video texts, while the number of participant’s column specifies the number and percentage of test takers who mentioned each aspect, regardless of how many comments on each aspect they made.

TYPE OF VISUAL	CATEGORY	ASPECT	EXAMPLES	NO OF COMMENTS	NO OF TEST
VIDEO	**Slides**	Textual cues	Words on PPT	17	10 (33%)
		Pictures and diagrams	A picture or a diagram related to the main topic.	11	7 (23%)
	Speaker	Facial expressions	Speakers' lip movements, head, eyes towards the camera.	20	15 (50%)
		Body language	Hand movements and gestures.	9	6 (20%)
		Character	Speakers' appearance and engagement with the topic.	15	11 (37%)

Table 4.7: Helpful aspects of the Video texts as reported by the test takers.

As shown in the table above, the results of the qualitative analysis in Nvivo revealed two main dominating categories related to the helpful aspects of the video texts, which are the slides and the speaker categories. Starting with the *speaker*, the most frequently mentioned aspect was the features related to the speakers' face expressions and motions, like their lip movements and the way they are looking towards the camera (20 comments from 50% of the test takers). In addition, speakers' character seems also to attract test takers' attention with 15 comments from 37% of the test takers. Traits of the speakers' character mentioned by the test takers include the way the speakers present the topic, their personal appearance and also their accents. Moreover, test takers mentioned aspects related to the speakers' body language (9 comments by 20% of the test takers), these aspects include examples like how the speakers occasionally use their hands to physically describe the shape of an object, pointing to the slides at specific points, or doing some gestures related to what they said.

Test takers also mentioned helpful aspects related to the *slides* in the video texts. Textual cues like topic titles and some main points related to the topic of the lecture or the outline of some lectures were mentioned in 17 comments by 33% of the test takers. In addition to these cues, there were 11 comments from 23% of the test takers which referred to the helpfulness of some of the pictures and diagrams or tables provided in some video texts. It is worth mentioning that all the textual cues and the tables/ diagrams support were in most cases presented in the question page as well, or – in the case of pictures- they displayed a photo which linked to the general topic of the lecture with no specific relation to any of the test items.

Below are two examples of the viewing patterns in the video condition by two participants as shown by the heat maps (Figures 4.2 and 4.3). The display of these maps to the participants helped them to refresh their memories about the way they have visualised the screen during their listening to the visualised texts and as a result reporting more information about the aspects they looked at and then providing a reason for their behaviour. This technique was considerably helpful to both the researcher and the test takers as it helped to show the relative intensity of the aspect in focus by shading the areas with the highest amount of fixation with a hot colour (red), and shading the areas with lowest fixations with a cold colour (green). However, seeing the red shaded areas does not always imply that the object in focus is being useful to the participant. Sometimes test takers focus on specific areas because they are a source of confusion to them and they try to clarify that confusion by focusing more on it. In other cases, some test takers do focus their gaze on a specific aspect (e. g., the speaker), while thinking about something else, like focusing entirely on the audio input. McIntyre (2016) describes this as a "matter of covert attention" (p. 292), which is basically a change of mental focus without moving the eye from the visual monitor. Fortunately, that can be clarified by the test takers themselves when they explain the reasons behind their viewing patterns of each visual text.

The first screenshot presented below (Figure 4.2) reveals that test taker 7 had a viewing pattern that focused on both aspects, the speaker and the slides, with slightly more focus on the speaker as its red shade appears bigger that the shade on the side of the slide, while the second screenshot reveals that test taker 3 (Figure 4.3) had general viewing over the screen with major focus on the speaker. These interpretations alone can be deemed superficial if not confirmed by the test takers' own reports about the reasons of their behaviours.

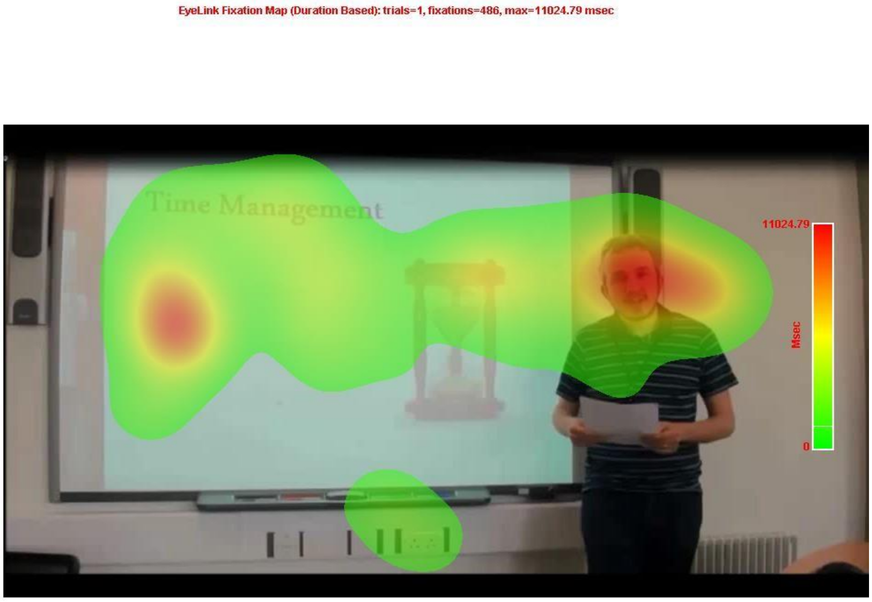

EyeLink Fixation Map (Duration Based): trials=1, fixations=486, max=11024.79 msec

Figure 4.2: Heat map shows the viewing behaviour of test taker 7 in video condition (Social Science).

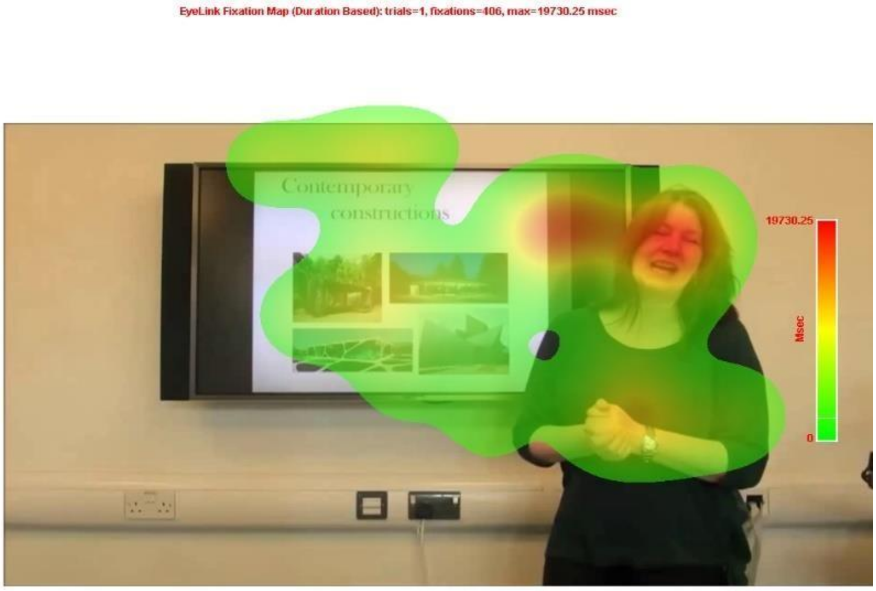


Figure 4.3: Heat map shows the viewing patterns of test taker 3 in video condition (Architecture).

In general, it seems that the test takers in this study focused their attention on speaker related aspects (44 comments in total) more than on slides related aspects (28 comments in total). The reasons of perceiving these aspects as helpful and useful are presented in the following section.

Reasons for finding video texts helpful:

Table 4.8 summarizes the test takers' reasons for focusing on specific aspects on the screen while watching the video texts. These reasons reflect their perceptions stated in Table 4.7 above about the useful aspects that test takers found in video texts and are grouped into major categories. Furthermore, they stated some other reasons that are generally linked to the video condition and not associated with any of the aspects they referred to in earlier incidents. In other words, they provided major reflections on how they felt regarding watching a video text. Accordingly, these reasons were grouped in three main categories (slides, speaker, and the need for visuals) in order to come up with a comprehensive result.

TYPE OF VISUAL	CATEGORY	REASON	NO OF COMMENTS	NO OF TEST TAKERS
VIDEO	**Slides**	Words confirm the audio input and help comprehension.	27	12 (40%)
		Pictures and diagrams summarize main idea.	22	11 (37%)
		Facilitates note taking.	12	7 (23%)
	Speaker	Lip movements, eye contact with the screen, face expressions help comprehension.	12	9 (30%)
		Gestures and body language help link spoken information to the slides.	12	8 (27%)
		Help to stay focused.	25	17 (57%)
	Need for visuals	Boosts comprehension and helps to remember information.	30	16 (53%)
		Reflects real situations Like real lectures.	24	15 (50%)

Table 4.8: Reasons of finding the video texts helpful as reported by the test takers.

While test takers were asked first about the aspects they found helpful, they were then prompted to explain the reasons which led them to perceive these aspects as useful. One of the most frequently expressed reasons was related to the general aspect of the *Need for visuals*, in which test takers explained factors that are not related directly to the speaker nor to the slides, but rather to the whole experience of watching a video text while they are taking the test. For this reason, it was preferred to assign a separate category for them. The Need for visuals seemed to be a very important reason that attracted the test takers to the screen during watching the video texts and considering them to be helpful. More than half of the test takers (53%, with 30 comments) indicated that seeing the whole environment of the lecture helps them to comprehend the audio input in a better way because it gives them the feeling of a real lecture, and that also helps them to remember more information, so they do not need to take lots of notes and that for them resulted in increasing the amount of attention to the visual spoken text. The following examples show how the test takers expressed this reason:

> *I found myself memorizing the information faster and I can also recall the information easier when I see a video* (Participant 4/ Health V)
>
> *If the pictures are moving, I mean if it is a video I will remember more* (Participant 10/ Zoology V)
>
> *I think with video I can remember more…I was able to remember more from the video. I can still remember it now*. (Participant 18/ Sports V)
>
> *When I was watching the video, I think I was able to store more information in my memory. So it was not important to take notes with the video compared to seeing nothing* (Participant 23/ Social Science V)
>
> *It was easier to me to see and listen to the information at the same time* (Participant 25/ Social Science V)
>
> *I found it nice for me to look at the whole environment, the lecturer and the PPT. That actually helps me to keep focused and does not let my mind wander away*. (Participant 3/ Architecture V)

> *I relied on my memory most. Especially with the videos, I felt that I do not need to take notes. I depended on my memory because the information stuck into my memory and it was easy for me to remember it*. (Participant 19/ Health V)

> *That is why I prefer the video because I concentrate on the information and I don't need to take lots of notes*. (Participant 5/ Sports V)

> *It is because I like the topic, and when it was presented in this way* (video) *it became more interesting. I was like, oh this is interesting I did not know there is a way of displaying things at the supermarket. It is something that is really close to me and that is why maybe I kept looking at the screen. I think that helps to comprehend more and be more engaged with the topic*. (Participant 8/ Business V)

> *I need to see the face. I am not sure whether the speaker's articulation or the movement of their mouths help me to understand and increase my listening comprehension, but I need to see their faces. I feel comforted that way*. (Participant 14/ Social Science V)

> *Whenever I need to understand something, I need to see it. This is how it works for me.* (Participant 30/ Health V)

There was also a significant percentage of the test takers (50%) who referred to videos as being helpful because they reflect the real situation of lectures in which students can see the lecturer and the slides related to the topic of the lecture. Below are some examples from the test takers:

> *Videos give me the feeling of interaction, which may help more than just pictures alone, and that makes me more focused and relaxed I guess…because I think that is how I will feel normally when I am interested in listening to a topic.* (Participant 11/ Business V)

> *I think my understanding improve a lot when I can see both the speaker and the whole surrounding environment, because for me this will be just like normal lectures. (*Participant 12/ Sports V)

> *Basically videos look more realistic. More normal like ordinary lectures. So if I have to choose between these two types* (i.e., video and still photos) *I will choose video. Much better for me.* (Participant 15/ Architecture V)

> *I think the video one is the more normal one for me, especially when you see the visual image moving naturally. This can help you to understand it more… I find it easier to follow the speakers' speech. It is normal like in a lecture.* (Participant 18/ Architecture V)

> *This video is easier than the still photo one, because it got the speaker moving, and it got some information on the slides. It makes it easier to follow the slides when the speaker is moving naturally in the video… it is just much more helpful to me. It feels much better and more normal than watching pictures that change at certain intervals.* (Participant 2/ Business V)

> *Actually I preferred the video one, because I saw more interaction from the lady, which got me more attracted into thinking and focusing more… I found it nice for me to look at someone speaks to me like that…it is more realistic.* (Participant 3/ Sports V)

> *Yes, it feels like I am in a real lecture and it is comfortable to look at the… presenter, because… this is a situation that I know about, it is like being in a classroom with a teacher.* (Participant 8/ Business V)

> *It is kind of attractive to see the situation. And I think with the development of technology teachers are now using the PPT where they have the outline of the lecture, or the main purpose of doing something by using pictures or direct words, so I think that makes me learn more from the screen here, just like what I learn from an actual lecture.* (Participant 9/ Sports V).

Beside reasons related to the need for visuals, test takers also referred frequently to reasons related to the slides and the speakers themselves. Regarding the slides, as test takers had commented on each visualised text individually, the reasons reported here are related to the slides that accompanied the video texts only. It is surprising that some test takers perceived the slides with the video texts in a different way from the slides with the still photo texts. Here is an example from one of the test takers summarising her experience regarding this issue:

> *In the video texts it is easier for me to follow what is on the screen. I found some differences between them. The one which included presentation and the speaker is moving naturally with the slides available it was much easier than the one with the still photos. I don't know, but the slides feel different when they are in the video mode.*
>
> *I can follow them easily*. (Participant 2)

Comments on the slides with the still photo texts are reported separately in Table 4.10.

Regarding the video texts, 40% of the test takers mentioned that the textual information on the slides helped them to understand the topic in a better way and memorise more information. They also explained that textual information boosted their confidence while listening, as the words on the slides confirmed the information they received from the audio input. Some representative examples are listed below:

> *I think seeing the words on the slides made it easier to understand it. It is like I am sure about the word when I see it*. (Participant 10/ Health V)

> *The main point for me and most helpful is to see the written information. Even though there was not a lot of information, but it gives you like an outline of what she wants to say about the topic*. (Participant 12/ Sports V)

> *Written words helped me to understand more, so I can get the information not only from the listening but also from the words on the screen. This makes me feel more confident. I think these few words make the lecture more organised*. (Participant 18/ Sports V)

> *When there are some notes on the power point it helps to understand the listening*. (Participant 20/ Social Science V)

> *I prefer to see written words, because I think these words are key words in the presentation, so I can focus more on them. It is helpful to understand in general like*

> *when there is a title, and that as a result helps me to understand the details in a better way.* (Participant 23/ Business V)

> *Written words are more helpful. When I see a word I can remember it easily, or take as a note. And also I feel more confident about it.* (Participant 25/ Social Science V)

> *I think the writing form is more helpful, because I can find precise words on the slides.*
> *It makes it easier and saves time of thinking about the word*. (Participant 6/ Health V)
> *Words on the slides are really helpful, because they give me an instant message of important ideas.* (Participant 9/ Architecture V)

There were also 22 comments from more than third of the test takers (37%) about the usefulness of providing pictures and diagrams in the slides that appeared in the video texts. Test takers explained that these pictures can give them an instant clue about the general idea of the topic. Some of these perceptions are explained in the following examples:

> *I think pictures work more for me, better than words, because pictures just give the whole idea of what the speaker is talking about, and I don't waste time on reading and understanding the words.* (Participant 10/ Health V)

> *Pictures and graphs are maybe very vivid, especially for the non-native speakers like us. I always remember pictures faster and easier than the words. And sometimes I don't even need to take notes because I can remember all the things on it.* (Participant 11/ Business V)

> *At the very beginning I didn't know what the word Stranding in the title means, so the first picture reminded me of the meaning of stranding.* (Participant 13/ Zoology V)

> *I think the pictures are more vivid and I can understand it without explanation.* (Participant 19/ Health V)

> *In pictures and diagrams, they give very important information to me, and you will get it quickly without much thinking. That will be more helpful, they just give me a summary of the lecture. I don't need to read a lot of words because when I read and*

listen at the same time that distract me and also I will get tired quickly. (Participant 22/ Health V)

I usually prefer a picture much more than written words because if the words are difficult they wouldn't help at all. But I can of course understand anything shown in a picture. It is easy because it doesn't involve any additional language thinking besides the listening. (Participant 25/ Business V)

The last reason related to the usefulness of the slides in the video condition is that some test takers (23%) found that seeing the slides during listening to the text helped them to take more notes because they can refer beck to the slide and that gave them self-assurance that what they recorded in the notes sheet is correct. It should be noted that the information presented in all the slides did not provide specific ready to use answers to any of the test items. Rather they simply related to the main idea of the topic or provided outline of basic points in the lecture that are already used in the questions screen of that text. Below some comments from test takers regarding this point:

I like that we can see the slides. It gives me idea what to write, and I think this is important for answer. (Participant 16/ Health V)

It important to me to take notes when I listen... so when I see the writing in the screen that helps me a lot. It really the best part in the video. (Participant 17/ Social Science V)

When I have listening test I like to look always at the answer sheet to find a clue about what they say. But here the slide makes it easy for me and that helped me to write so many good notes. (Participant 27/ Zoology V)

It made me more sure what I note down is right. (Participant 11/ Social Science V)

Beside the reasons related to the slides, it seems that test takers have largely considered several helpful reasons related to seeing the speaker in the video text. A significant number of test takers (75%) found that seeing the speaker moving and acting naturally in the video text helped them to focus more on the audio input. Here are some examples from the test takers' comments:

> *I just couldn't stop looking at the person talking. Clearly I looked at the speaker more than the slides because I find looking at him is more attractive to my attention. I can listen better* (Participant 13/ Zoology V)

> *Some speakers are very helpful like the one here, she is really interesting and explains the topic in a good way and that drives me to look at her.* (Participant 17/ Sports V)

> *I think whenever I can see the speaker moving I will look at it… It is more…engaging* (Participant 18/ Architecture V)

> *The speaker's way of talking is very clear so I keep looking at her, she makes me focused on her speech… generally seeing the speaker is far more important to me than seeing the slides. I can do without the slides but not without the speaker.* (Participant 19/ Health V)

> *I think it is very helpful to see the person speaking…I just like to see the speaker in front of me. It feels natural and gets me to listen in a better way.* (Participant 21/ Architecture V)

> *In fact, how she interacted made me like more concentrated at the topic.* (Participant 3/ Sports V)

> *It is just my normal behaviour to look at the speaker when they are talking in order to feel engaged with what they are saying*. (Participant 4/ Zoology V)

> *I like the way she was interacting with the topic…it really helps a lot and kept me engaged.* (participant 8/ Business V)

> *I think it is normal to see who is talking to you, right? So whenever the he is moving I just don't like to look anywhere else but on him… I think this is more helpful to understand the lecture*. (Participant 30/ Zoology V)

In addition to reasons related to seeing the speaker in general, a number of test takers were more detailed and pointed out to some specific reasons. For instance, 30% of the test takers revealed that factors like the speaker's lip movement and eye contact or their facial expression are very helpful to comprehend the spoken input. They found these factors to play a complementary role to the audio input and help them to understand the text in a better way, below are some of these comments:

> *Looking at the man speaking here…with moving video I can track his lip so I think I would understand o lot more.* (Participant 10/ Zoology V)

> *Yes, here I prefer speakers like this who look at me and…what is it called? eye contact! yes make eye contact with me rather than focusing on the paper only.* (Participant 8/ Business V)

> *Of course I find the speaker helpful because I can see her lips moving when she speaks. It is important to me. Even in normal conversations I look at the person's lips when talking.* (Participant 12/ Architecture V)

> *Because we are non-native speakers of English language, I think the movement of the lips is very important to us. It helps me to understand more, so seeing the women here speaking in a clear way was very helpful to me.* (Participant 17/ Business V)

> *The man's lip movement helps me stay focused.* (Participant 28/ Social Science V)

> *I think her lips moving in a clear manner and it helps me a lot to understand. And I also find it helpful to look at her face, when talking about something not good for our health she looks a bit sad.* (Participant 24/ Health V)

> *I like to see the person who is speaking. I think it is helpful to see some… facial… expressions. This helps sometimes to understand how the person moves from one point to the other. I think it is interesting*. (Participant 23/ Sports V)

There were also eight test takers (27%) who found the speakers' body language and the gestures they made in the video texts to be helpful in linking the audio input to what is presented in the slides and consequently increasing their understanding of the topic. They also revealed that they find speakers who use body language during their talk to be more helpful in keeping them focused. The following examples reflect this concept:

> *Sometimes their body language, yeah… it could help specially when he points to the slides when he is talking about the points he discusses. (*Participant 2/ Social Science V*)*

> *Well... the body gestures… even at times when she is raising her hands or looking at the camera. I feel she is communicating with me. So it helps a lot. (*Participant 29/ Architecture V*)*

> *Sometimes body language is very important. First it keeps me focused to the speaker and it helps me to connect what I hear in the listening to the way the woman here is talking. I think sometimes I can even understand a general thing from her body language when I find some difficulties with the listening itself. (*Participant 25/ Business V)

> *I like the way she was interacting with the topic using hand movements that really kept me engaged…see, when the body movement goes on time with the important information, that helps me to remember the information.* (Participant 8/ Business V)

In sum, helpful aspects and reasons perceived by test takers with the video texts are mostly revolving around the role of the speaker as an active figure. The slides were also perceived as helpful and, together with the speaker, they help to reflect real-life academic lectures.

In addition to the helpful aspects and reasons that a number of test takers found in video texts, some has also caught few helpful aspects in still photo texts as shown in Table 4.9 below.

Helpful aspects of the still photo texts:

Table 4.9 summarizes the main helpful aspects that the test takers found in the still photo texts. The two reported aspects are related to the general design of the lecture including the display mode. There were not any comments in this respect related to the speaker, and this suggests that the test takers probably did not find the presence of the speaker's figure in a still photo mode to be helpful to their comprehension.

TYPE OF VISUAL	CATEGORY	ASPECT	EXAMPLES	NO OF COMMENTS	NO OF TEST TAKERS
STILL PHOTOS	**Lecture design**	The slides	Words and diagrams.	18	13 (43%)

		Pausing time	Still photos change at certain intervals.	7	5 (17%)

Table 4.9: Helpful aspects of the still photo texts as reported by the test takers.

As the table reveals, test takers expressed only two main helpful aspects related to still photo texts. Less than half of the test takers (43%) referred to the helpful role of the slides in some still photo texts. They found the slides and their content of textual information and diagrams or pictures to be particularly more helpful than seeing the speaker in a still photo text. It is important here to mention that because these results are reported according to the helpfulness level not the text level, these results could possibly be affected by the content of some specific texts. In other words, the usefulness of the slides that test takers reported here is probably not merely because the slides were in a still photo mode, instead because of the content of the slides themselves, regardless of being in a still photo or video text, found to be useful to them. The following comment gives a clear example about this point:

> *It sometimes doesn't matter if it is a video or pictures only in this matter. When the slides give you information you will look at it regardless of whether it was in video or picture.* (Participant 3/ Health SP)

The second aspect is exclusively related to the still photo texts and the way they are displayed. 17% of the test takers found that a helpful feature of the still photo texts is the regular pausing of the pictures that only change at certain intervals. This design of still photo texts seems to suite some of the test takers more than the moving video design. The reasons are explained below (Table 4.10) as stated by the test takers.

Reasons for finding still photo texts helpful:

In addition to the two helpful aspects of still photo texts, the test takers gave a more detailed account to the reasons of the usefulness of these texts as revealed by the summary in the following table.

TYPE OF VISUAL	CATEGORY	REASON	NO OF COMMENTS	NO OF TEST TAKERS
STILL PHOTOS	**No distraction from the speaker**	Speaker is not moving helps to focus more on the slides and the audio.	7	5 (17%)
	Summarize the whole situation	Gives a good idea about the context of the lecture.	11	8 (27%)
	Helps to take notes	Gives more time to write notes.	8	7 (23%)
	Easy to ignore	When everything is still, it is easy to ignore the screen if the test takers decided.	4	4 (13%)

Table 4.10: Reasons of finding the still photo texts helpful as reported by the test takers.

As the table above reveals, it seems that the most important reason for perceiving visuals to be helpful is that still photo texts gave them a clear idea about the environment and the context of the lecture to which they are going to listen and that, for them, serves as a summary which helps them to focus more on listening and understand the audio input in a better way. This reason was acknowledged by 27% of the test takers, which represents the largest category in reasons of helpfulness of still photo texts as suggested by Table 4.10.

The following quotes clearly explain this reason:

> *Yes, this type of text helps a lot, because it is like kind of summarizes important information about the lecture.* (Participant 1/ Sports SP)

> *With the still pictures I look at it and I know that this is still, so it summarizes the whole situation to me. Like in this Zoology text here, I only looked at the screen, a man speaker and slides with a picture so I knew what is it about. Clear.* (Participant 17/ Zoology SP)

> *I prefer the still photos. It makes me more focused on the listening because it gives me an idea of the thing they are going to talk about, how the speaker looks like and if there are any other information like the slide here. That helps a lot in my opinion.* (Participant 24/ Social Science SP)

> *One glance at the screen can give me all the information I need. It has got every thing important like who is speaking and what is on the slides.* (Participant 3/ Business SP)

Another reason reported by the test takers is that still photo texts help them to take more advantage from the slides and the audio input, because the speaker is not moving. In a total of 7 comments, those test takers claimed that when the speaker is moving like in the video texts this in fact distracts their attention because they are automatically attracted to looking at the speaker and they forget to focus on the audio information. Therefore, they find the still photo texts more helpful because the speaker is not moving. Interestingly, this reason seems to contradict what other test takers reported regarding the helpfulness of the slides with the video texts, as they found that the movement of the speaker as he/she is referring to the slides from time to time had helped them to focus more on the slides. As for the still photo texts, some of the comments that describe this reason are listed below:

> *I think pictures are a bit better than video because basically during the picture text the speaker won't move so he will not distract me.* (Participant 16/ Sports SP)

> *Pictures are little better. Because they are fixes, and they don't move compared with the video so I will not be distracted… I can look at the screen without interruption from the movement of the man here.* (Participant 22/ Social Science SP)

> *Pictures like this one are the best for me mainly because I don't see the body movement of the speaker and I can still see the slides without the distraction of the body movement.* (Participant 24/ Social Science SP)

> *I find this type is really good actually. Because you know… it is easy to look at the information on the slide which are the most important thing to me, and not to look at the speaker who is the source of confusion for me.* (Participant 29/ Business SP)

Related to the previous reason (speaker is not moving), 13 % of the test takers claimed that when the speaker is not moving, it is easier for them to ignore the whole visual input and to focus only on the audio input. Unsurprisingly, this reason was mostly reported by those who found the audio-only mode to be the most helpful, hence they felt that the still photo mode, unlike the video one, can be easily controlled by ignoring it and treating the text simply as an audio-only. The following comments explain this reason:

> *Here when the speaker is not moving, I can focus on the listening information without any other thing that… makes me less focus, like the speaker or anything in this screen.* (Participant 1/ Social Science SP)

> *It is easy to ignore it if I decided to.* (Participant 20/ Zoology SP)

> *I didn't like it when the speaker was moving because I felt I am obliged to look and if I didn't look I might miss something. But here when the pictures are still, or not moving, it is more easy to ignore it all together and just listen to him.* (Participant 13/
> Social Science SP)

There were also 8 comments (from 23% of the test takers) who claimed that still photo texts were helpful because this mode helps them to write down more notes. They explained that this feature allows them to look briefly whenever the picture changes and gives them a general idea about what new information this picture can add to them. Then they can relate this new visual information to the audio input and write good quality notes. Here are some quotes that explain this reason:

> *I actually prefer this still photo type, because I can still see the slides which can help me understand some more information and at the same time take notes. I feel more*

> *relaxed because I know here I won't miss anything new from the picture until it changes to the next one.* (Participant 24/ Sports SP)

> *When the test is in still photo mode it makes it easier for me to have some time focusing on the information in the slide, and then I can write notes if I want to without feeling that I'm missing something on the screen.* (Participant 6/ Social Science SP)

> *I think I need to see the visual first to get something in my mind, then I can take notes. and this mode* (still photos) *gave me this advantage. I mean gives me the time to look first, like scanning the screen, then write the notes.* (Participant 9/ Health SP)

Helpful aspects and reasons for video and still photo texts

Inspecting the aspects and the reasons for the helpfulness of video texts and still photo texts as perceived by the test takers (Tables 4.7, 4.8, 4.9, 4.10), one can observe an interesting pattern. It can be noticed that the helpful aspects and reasons related to the still photo texts did not include any reference to the speaker's role. There was only one reference in the reasons of finding still photo texts helpful, which basically refers to the inactive appearance of the speaker's figure in this type of texts. All other aspects and reasons were in fact related to the general design of the lecture or to the slides, but not to the speaker. Comparing this to the aspects and reasons reported by the test takers on the video mode, it is clear that the speaker plays a very important role when it comes to video texts. However, this important role of the speaker in the video texts is accompanied by other factors like the slides and the whole design of the video texts, which are aspects reported in the still photo texts as well.

These findings suggest that the appearance of the speakers is particularly helpful when it is in the video mode, where the test takers can relate the speakers' movement, gestures, body language or facial expressions to what is presented in the slides and to the stream of information in the audio input. In other words, the results indicate that when the visual information in the video texts is semantically congruent with the audio information (audio information is synchronised with visual information) and presents a whole dynamic content (speaker is interacting with visual and audio information) that is rhetorically effective (resembles real live situations), it proved to be more helpful than the visual information in the still photo texts which is semantically neutral, liminal in its content and rhetorically neutral or ineffective.

More specifically, test takers in many instances described the benefits of being able to see the non-verbal, kinesic behaviour of the speakers in video texts, in addition to the physical context of the lecture. These factors were very important to those participants in assisting them to create meaning from the audio input, and by that it can be argued that providing the test takers with the basic natural setting of the spoken text, and the opportunity to see the speaker can positively affect the way they process the spoken input, and help them to get the most out of that text, therefore this design of including video texts should be considered by test makers when designing a listening comprehension test. It has been claimed by some researchers that the use of video texts in listening test can affect the construct validity of the test because it would be considered as a construct irrelevant variable when it comes to assessing the listening ability of test takers and therefore should not be included in the test (Buck, 2001; Coniam, 2002). However, it seems contradicting that Buck
also contended that if the goal of the test designer is to evaluate test takers' ability in particular modality, it is central that the unique characteristics of that modality should be included in the test (2001). In a similar way, Rost (2013) argued that it is very important to pay attention and include features of spoken language that are unique to listening if the goal is testing listening ability of test takers. This, in Rost's opinion, would make test designers more confident with the construct validity of the test than if these features were not included. As proved by the results of the data analysis in this study that the inclusion of video texts, with all its features, was essential to many test takers in comprehending the spoken texts, and they confirmed that this visual information was an integral part of the way they processed the input. Therefore, excluding this type of information from the listening test would raise the issue of construct underrepresentation, which in turn would be a clear threat to the construct validity of the test (Messick, 1996). It is reasonable therefore to argue that the use of video texts in L2 academic listening comprehension tests support the construct validity of the test, as it provides an environment that reflects the characteristics of the academic language domain.

Another finding related to perceiving video texts as helpful was that these texts reflect real life situation as stated by half of the participants (50%) in the Need for visual category. For instance, when Participant 8 stated that "*it feels like I am in a real lecture… this is a situation that I know about, it is like being in a classroom with a teacher, and that makes me interact naturally with it"*, she referred to the fact that the video text had generated an atmosphere that was not like a test situation, instead it gave her the feeling that she is in a real lecture, and she behaved according to that in processing and answering the test items. This as a result had potentially minimized the test-specific behaviour, which is normally a dominant trait in most test-situations. It is suggested that when there is a prove of test-wise strategies among test takers, it would be a threat to the cognitive validity of the text (Field, 2010). Therefore, this finding can be possibly used to support the notion that video texts can improve the cognitive validity of the text by reducing the test specific behaviour among test takers.

Another very important feature that test takers found helpful in video as well as still photo texts is the use of slides. Most test takers attended to the slides at some point during their watching as proved by the heat maps. For many of them, they reported that they found the slides useful and supported their comprehension of the text either through the textual or pictorial information or both. These types of information displayed in the slides were used in different ways by different listeners. For instance, for some of them, they associated the textual or pictorial information with what they already have in their memories. In other words, the slides worked to activate specific schemata related to the topic they listen to, and this activation of the suitable schemata is vital to actual comprehension (Rost, 2013). For others, the information on the slides were used to evaluate and confirm the spoken input, and that is a sign of using some metacognitive assessment strategies like monitoring and problem-solving strategies (Vandergrift and Tafaghodtari, 2010), which are key strategies that help listeners to comprehend the message properly.

Regarding the still photo texts, some test takers found the fact that the speaker is not moving to be helpful, and the heat maps revealed how they focused more on the slides' zone than on the speakers' zone. That is also linked to another reason for finding still photo texts helpful, and that was because test takers found them easy to ignore. It could be looked at as a cause and effect relationship. Because the speaker is not moving, the still photo texts are easy to ignore. It is possible that those participants who preferred the still photos over the moving ones (video texts), have some difficulties in handling more than one modality at the same time (audio and visual modalities), and therefore preferred to ignore the screen and focus only on a single modality, that is audio. Those test takers mostly preferred the audio condition over the two visual condition in general, and preferred the still photo condition over the video one.

Still photo texts were also perceived as helpful by some test takers because of its design that changes the still pictures at certain intervals. This design provided those test takers with more time to write notes after looking at every new photo and inspecting the type of information it carries. As revealed by the data of the cued retrospective report, some test takers who found this aspect useful rendered it to the fact that each photo will take some time to change to the next one, so they felt more comfortable to write notes without the "distraction" -as they described it- of the moving visuals in the case of the video texts. In fact, note taking seems to be an important trait for academic lectures' listening (Flowerdew, 1994), and the capability to take notes while listening to a lecture found to be advantageous for L2 listeners (Bloomfield et al., 2010). However, while some test takers found this aspect to be helpful, others reported that they preferred a "*less distracting design*" that can minimize the need for taking lots of notes. This aspect along with a number of other factors that test takers found distracting are discussed in the following section.

Distracting aspects of video and still photo texts

Beside the helpful aspects that the test takers found in video and still photo texts, a number of them also reported some distracting aspects in these visual texts. It is also important to mention this as the data was analysed according to the visual type level (video or still photo), not to the individual texts level, it is possible that some test takers could have found a specific aspect to be useful in one video text (e.g., the speakers' appearance and movement) or in a still photo text (pictures and diagrams), and distracting in the other.

Distracting aspects of video texts:

Table 4.11 summarizes the main distracting aspects that test takers found in the video texts.

TYPE OF VISUAL	CTEGORY	ASPECT	EXAMPLE.	NO OF COMMENTS	NO OF TEST TAKERS
VIDEO	**Nature of video texts**	Concurrent audio/video stimuli	Listening, watching, and taking notes at the same time.	6	4 (13%)
	Speaker	Speakers' movement	Eye contact and body language.	9	5 (17%)

Table 4.11: Distracting aspects of the video texts as reported by the test takers.

As shown in the table above, there were mainly two distracting aspects related to video texts as reported by the test takers. More specifically, 13% of the test takers claimed that the display of video lectures is found to be more demanding than the still photo lectures and the audio lectures. In other words, during the video texts they felt that they have to listen, watch and take notes simultaneously to make sure that they did not miss any part of the lecture, and that process proved to be distracting and sometimes overwhelming to them. The other distracting aspect is related to the speaker. 17% of the test takers found the movement of the speaker in video mode texts to be distracting. The speakers' movement here refers to a range of aspects starting from simple movements as little as eye contact to physical movements of body language or walking towards the slide to point to a specific point in it. This aspect was mentioned by some test takers as a helpful aspect in one video text (like Participant 8/ Business video) and distracting in another video text (Participant 8/ Social Science video). With the aim of explicating these results, it is important to inspect the reasons that led test takers to consider some of the visual aspects in video texts as distracting.

Reasons for finding video texts distracting:

The two distracting aspects reported by some test takers above in Table 4.11 about the video texts had generated more details about why did those participants think that these aspects were distracting. On the whole, two main reason were highlighted by test takers to describe how the video texts were distracting to them. These reasons are summarized in Table 4.12 below.

TYPE OF VISUAL	CTEGORY	REASON	NO OF COMMENTS	NO OF TEST TAKERS
VIDEO	**Excessive focus on the video.**	The movements on the video attracts test takers' attention and interferes with the process of note taking and leads them to miss parts from the listening.	9	5 (17%)
	New type of listening.	Test takers are not prepared or used to watch a video during listening tests.	11	9 (30%)

Table 4.12: Reasons of finding the video texts distracting as reported by the test takers.

The most frequently mentioned reason out of the two is that test takers were not prepared to have a listening comprehension test that is supported with video. 30 percent of the participants (all Chinese) explained that this new condition was not available to them when they took the English language test preparation courses. In other words, the nature of the video condition contradicts with the strategies they learnt in that kind of courses. Some examples from the test takers' verbal reports explain this reason:

> *When I took the IELTS course to prepare for the test I didn't take any with video like that. This I think confuses me more and I cannot focus to do the tactics I learnt in that course. It does not help me.* (Participant 20/ Zoology V)

> *Back in China we did not do that video type. It is totally new to me.* (Participant 15/ Social Science V)

> *I think it looks more like a lecture not a test… that makes me forget about the test questions and not concentrating on how to answer the questions. Because you know we learnt about how to answer test questions, but this video makes me forget. I prefer the way I was taught how to do the test. This one here is like watching TV programme or something.* (Participant 1/ Health V)

> *This video let you think about more… different things, but not about the way to answer the test. I was very good for using… tricks… or techniques to answer the test when I was back in China taking the preparatory course for English. But here, this doesn't look like that at all. It makes it more confusing.* (Participant 22/ Health V)

The second reason revealed by the test takers shaded the light on the nature of the video texts that required the concurrent use of video-audio stimuli and the continuous movement of the speaker. These elements caused some test takers to be overly attracted to the screen and that as a result led them to miss parts from the audio-input and forgetting to take notes as reported by 17% of the participants. Some of their comments are listed below:

> *I just didn't know what to focus on with the visual texts, the video I mean. If the thing or the person is moving I can't help but to focus on it and follow it. Then I discover I forgot to focus on the listening. You know it is not easy doing two things at the same time. (*Participant 16/ Zoology V*)*

> *The video for me is more distracting because I try to follow it and when I look down to write notes and then look back at the screen I try to connect...or link what I'm seeing with the last thing I saw in the screen, and also to listen at the same time. All this gets me completely distracted. (*Participant 17/ Social Science V*)*

> *When the person is moving my focus shifts from the listening to their movement. So instead of focusing on the listening passage itself I find myself focusing more on the visual. It sometimes gets me very distracted.* (Participant 24/ Zoology V)

> *Video distracts my attention because when I watch it I miss many parts of the meaning and content of the talking.* (Participant 20/ Business V)

While test takers reported only two main distracting reason with regard to video texts, the results of still photo texts revealed more distracting reasons in both their categories and numbers of test takers who commented on those reasons. Concerning the distracting aspects of still photo texts, the result was similar to the video texts with two aspects describing the areas of distraction to the test takers.

Distracting aspects of still photo texts:

Test takers assigned two aspects related to still photo texts as distracting. These aspects are described in Table 4.13 below.

TYPE OF VISUAL	CATEGORY	ASPECT	EXAMPLE	NO OF COMMENTS	NO OF TEST TAKERS
STILL PHOTOS	**Lecture design**	Pausing time	The still pictures that change at certain intervals.	12	10 (33%)
	Nature of still photo texts	Concurrent audio/visual stimuli	Listening, watching, and taking notes at the same time.	8	8 (27%)

Table 4.13: Distracting aspects of the still photo texts as reported by the test takers.

Two major aspects were assigned by test takers as distracting in still photo texts. As shown in the table above, the first aspect is related to the general design of the still photo texts and the way this type of texts is displayed. Around third of the test takers (33%) described the fact that still photos in this text type, which change at regular intervals, are found to be distracting. Interestingly, this same aspect was perceived by some test takers as a helpful aspect as previously presented in Table 4.9. With regard to the second aspect, 8 test takers (27%) reported that having to listen, watch, and take notes at the same time is found to be distracting to their attention. This aspect has in fact been reported as distracting with the video texts as well. It seems that the visual texts in general, regardless of being video or still photo do sometimes have the same effect on test takers when they compare these visual texts to the audio-only texts.

In order to explore these results further and find out the rationale behind them, it is important to examine the reasons for considering still photo texts as distracting by some test takers.

Reasons for finding still photo texts distracting:

Test takers were asked during the cued retrospective reports to demonstrate their reasons behind perceiving still photo texts as distracting. Data analysis revealed several reasons as summarized in Table 4.14.

TYPE OF VISUAL	CATEGORY	REASON	NO OF COMMENTS	NO OF TEST TAKERS
STILL PHOTOS	**frequent pauses**	The frequent pauses distract the attention	15	11 (37%)

	Speaker is not moving	No clue from the speaker to help comprehension.	15	10 (33%)
	No connection between audio and pictures	No real time reflection of what the speaker says and what goes on the screen.	7	5 (17%)
	Excessive focus on the visual aspect of the still photo text.	The visual attracts test takers and distracts them from taking notes and leads them to miss parts from the listening.	14	7 (23%)
	New type of listening	Test takers are not prepared to watch still photos during listening tests.	13	10 (33%)

Table 4.14: Reasons of finding the still photo texts distracting as reported by the test takers.

According to Table 4.14, five main reasons were found in the data with regard to the visual condition of still photo texts that test takers viewed to be distracting. The most frequently mentioned reason, by more than third of the test takers (37%) is that the general design of the still photo texts was found to be distracting, mostly because the pictures change at regular intervals and test takers felt the need to look at each picture every time it changes. Despite the fact that the researcher has explained the different types of texts to the test takers before starting the test and showing them warming up texts that included the two types of visuals, some of the test takers did not perceive the still photo texts as a normal type but rather as a strange and distracting type. A sample of the comments below explain this reason more clearly:

> *I prefer not to look at the screen at all if it is pictures like this. Because I think with the pictures not moving like this, it feels like something wrong with the video. It is like the video doesn't work or jammed. that doesn't feel normal. Doesn't attract me*. (Participant 19/ Social Science SP)

> *This kind of test, I mean the pictures do not move as ordinary thing, and that makes me feel that I have to look each time I feel a change in the screen. That confuses me.* (Participant 1/ Sports SP)

> *The annoying thing to me is that I tried to focus on the listening itself, but whenever the screen moves to the other picture I look again. Unconsciously. That drives me crazy.* (Participant 10/ Social Science SP)

> *Each time the picture changes I had to look to see if there is something new in that picture. I prefer if there was only one picture from beginning to end instead of many pictures like these ones. This way is much better to me I think.* (Participant 29/ Business SP)

> *I did not like it. It is confusing and puzzling, I didn't know when to look, and when I look at the picture it dose not move, why? It should move it is better when moving.*
> *More normal, but this one here, no… it is not helping ... because that was not normal to me. Confusing me so much.* (Participant 27/ Social Science SP)

> *This one is difficult to me. I think because there are lots of pictures. And they change. If only there is one picture, or if it is movie* (i.e., Video) *that is more normal. Much better. But like this one is confusing.* (Participant 30/ Sports SP)

> *These still photos, well they weren't as useful to me as the videos…I mean here the same speaker appeared in the video text, but there she was moving her mouth and her hands, which was more normal and...genuine. But like this still pictures it was not realistic to me*. (Participant 4/ Sports SP)

The second reason why some test takers found the visual condition of the still photo texts distracting was because they got no clue from the non-moving speaker (body language and gestures) that may support their comprehension. Third of the test takers (33%) thought that when the speaker is not moving naturally like in video texts they feel more distracted as they keep looking at the speaker whenever the photo changes in the hope that they may spot a new supporting hint. Some quotes from the data are presented below:

> *This is not familiar…I couldn't see any help from the speaker in these still pictures. No lip movement or anything to help me understand more.* (Participant 10/ Sports SP)

> *I think if there was a slide only without the lecturer that will be more helpful to me. It is like I can evoke my background knowledge easier and faster. But the speaker is kind of useless or even distracting sometimes in the still photo design.* (Participant 4/ Social Science SP)

> *I wish if the speaker is moving naturally like in the other normal video. That was much better, because, you know, you can see many things like the face expressions and the hands moving and referring to the slides and so many other things which I think can be very helpful to me. I like to see the speaker moving…but here it is a bit difficult like this*. (Participant 29/ Health SP)

> *This type doesn't give any helpful hints for me*. *Like when I look at the speaker, nothing moves. She is just paused. I think I need to see the speaker moving like in the natural way.* (Participant 5/ Business SP)

In addition, 17% of the test takers explained that the nature of the still photo texts makes it difficult to synchronize the audio input with the visual information in the text. This reason in fact could be partially linked to the previous one (the speaker is not moving) since the difficulty of synchronizing audio and visual information is believed to be, to some extent, caused by the absence of the speakers' movement and body language, namely, when they refer to some clue in the slides while they are talking about a specific point related to it. Here are 4 of the 7 comments stated by the test takers:

> *There was a diagram at this point, but I don't know, I just couldn't follow it because I didn't know if the speaker was talking about it. It was kind of misleading when it first appeared. Didn't know when I should focus on it.* (Participant 9/ Business SP)

> *I think the moving pictures or photos like these may confuse my attention, and it is very hard to find related information according to the audio. I think if the man speaking here is referring to the PPT by his hands or so, it would be easier to link information correctly*. (Participant 11/ Zoology SP)

> *In this still photo mode, I couldn't follow the slides properly. Not like with the video one*. (Participant 2/ Zoology SP)

> *For me if there should be visual it should only be video, not like this one. It was just confusing in my openion. I cannot connect what this man says with what is in the slides.* (Participant 27/ Social Science SP)

Also, 7 of the test takers (23%) reported that the presence of visuals in the still photo texts is an important reason to consider this type of texts as distracting. In other words, the visual caused them to be excessively focused on the screen to the extent that it led them to oversight the content of the audio input and also not taking enough notes. Interestingly, this very reason was also reported in the video texts as the only reason for considering videos distracting in listening comprehension. It is worth mentioning that only 2 out of the 7 test takers who considered this reason to be distracting in the still photo texts have also reported that in the video texts. Furthermore, it is not a surprise to find out that those 2 test takers perceived the audio-only texts to be the most helpful and the still photo texts and the video texts to be generally distracting. Some quotes from the data are given below:

> *When I look at the screen maybe I will forget to take notes…I just looked at the pictures and the slides and I was thinking about how this lady looks… she is attractive… and then I was thinking of different things I remember a TV show presenter who speaks like her. At the end of the passage I realised that I missed a lot of what she said.* (Participant 16/ Sports SP)

> *I prefer to listen only. Because the visuals again here distract my attention …because when I watch I miss parts of the talk and I don't take good notes to answer the test questions.* (Participant 20/ Zoology SP)

> *Here I tried to ignore the picture, but I couldn't. It just gave me the feeling that I should look at the screen or I may lose something. Actually looking at the screen made me lose some of the listening.* (Participant 19/ Social Science SP)

> *That is not a good type. I looked here* (pointing to the screen) *too much and I didn't take notes. It confused me.* (Participant 28/ Zoology SP)

> *When I look at the pictures in the screen my mind wanders away.* (Participant 3/ Business SP)

Lastly, third of the test takers (33%) reported that the still photo condition represents a new type of listening test, because its design is not familiar to them in the case of testing. The same reason was reported also with the video condition as shown previously in Table
4.12. some quests from the participants' verbal data are presented.

> *I think this type was strange to me and I did not like it because… when I know that this is test, I am prepared to do test strategies, and I think we Chinese are very good in learning test strategies, like how to get the main idea and also how answer test questions with four choices and things like that. But here these pictures did not allow me to do that and it was looking like what I do in lecture, so I think I missed some strategies.* (Participant 1/ Social Science SP)

> *I'm not used to see just pictures in test of listening. It is the most difficult one to me. I was confused how to answer questions if I was not listening only. This situation is not like test situation.* (Participant 11/ Architecture SP)

> *It is fooling me. I mean it is kind of awkward. I was trained to do tests in a completely different way than this.* (Participant 15/ Business SP)

> *When I saw the screen at first I thought it is a video, but then takes me some time to realise it is a picture… still photos don't feel very comfortable. I just didn't like them.*

> *I'm not used to do a listening test and see pictures like this. It is not very normal to me. And I think that made me losing some techniques I know about tests.* (Participant 18/ Business SP)

> *The picture ones look a little bit weird, because they only capture a specific moment of the movement of the person and then move to another. So it is just something strange to me, not natural or familiar in listening test to see pictures like these. That's what I feel about it. Listening tests should only be sound because this is the way we trained for. At least this is what I think!* (Participant 21/ Business SP)

Distracting aspects and reasons for video and still photo texts.

The overall results as presented above (Tables 4.11, 4.12, 4.13, 4.14) revealed a number of distracting aspects in both video and still photo texts and showed how the test takers demonstrated their perceptions of these texts as distracting. In general, there were two distracting aspects for each condition. These aspects were partially similar in the video and still photo texts, as they shared one of the two aspects stated by the test takers. In both conditions, the participants found the nature of the visuals that required a concurrent attention to the audio/visual stimuli to be distracting. Sharing this aspect with both visual conditions might have an implication that those test takers who were distracted by the visual aspect in the video and still photo texts have in essence difficulties with multitasking in general. For this reason, they found it difficult and overwhelming to have more than one modality at the same time and couldn't divide their attention appropriately among these modalities, which resulted in a feeling of confusion. This factor was found in other studies as well. In Suvorov's (2013) study, a number of participants reported the same outcome as they were also affected negatively from the "information over load" found in the two types of videos used in his study. Interestingly, the other distracting aspects reported in the video and still photo texts were also reported as helpful by other test takers. For example, the Speaker aspect was found by few participants to be distracting with the video texts, while more participants found it to be rather helpful. Similarly, with the still photo texts, some participants stated that the lecture design that necessitate the change of photos at regular intervals which give some pausing time in the screen to be distracting and confusing. The same aspect was also reported as helpful by some participants. These differences in perceptions might be caused by the individual differences among participants. It seems likely that test takers have different abilities when they process the visual aspects in both video and still photo texts. Many researchers empirically found that individual listeners process spoken input in different ways (Buck, 1991, 2001; Goh and Hu, 2013; Rost, 2013). Although these studies inspected performance with audio only texts, their results are reflected in this study's findings as the test takers' descriptions of their processing of the texts differed as well. Several reasons might be discussed to count for these extensive differences. It might be again the case of different abilities in multitasking as explained earlier. For instance, those who found the presence of the speaker to be distracting in video texts might be the ones who have problems to distribute their attention between two or more modalities (e. g., listening, watching, writing notes), therefore the speaker's movements distracted them from focusing on listening only. Individual differences can affect listeners' perceptions about the visual conditions in many other ways like differences in their listening strategy use or even their learning style. However, the scope of this research does not extend to cover the issue of individual differences.

Regarding the reasons of perceiving the two visual conditions as distracting, it seems that the still photo texts had a greater distracting effect on test takers as they specified more detailed accounts for the reasons of finding this type of texts distracting compared to the video texts. There was a total of five main reasons for perceiving still photo texts as distracting when processing them, while with the video text only two major reason were stated by the test takers. Both of those reasons which are related to considering the video texts as distracting (the visual attracts test takers' attention and prevents them from taking notes, and that the use of visual is new to them and they are not used to it) were also stated with the still photo texts. The number of test takers who stated that the excessive focus on video texts led them to take less notes and miss parts of the spoken input was less than those who made the same comment with the still photo texts. This implies that the distracting impact of this issue was larger in its magnitude with the still photo texts, and that might be caused by the design of these texts. In other words, it could be possible that because the still photo texts change at regular intervals, with each change those test takers automatically look at the screen, and they spend longer time inspecting the new photo which might have led them to miss parts of the listening and not taking notes of what had been said. With the video condition on the other hand, it seems to be easier to ignore the whole visual input once the participant decides to do so, and not looking back to it if he/she felt distracted. For those participants, the visuals were mainly distracting because they prevented them from taking notes that might have helped them to answer the test questions, and that was possibly more important for them than watching the visual. This might mean that the visual materials did not help those participants to remember information, and that could be related to their ability to interpret and store audio-visual information.

The other distracting reason which was present in both video and still photo conditions is that the inclusion of visuals in listening tests is new and represented a surprising factor for some test takers. Those who were distracted by such factor claimed that they are not accustomed nor prepared to see visuals during listening tests, and that led to fail in applying the strategies they were trained to use in listening test situations. Those participants did not accept the visual because it contradicted with these test-wise strategies, and consequently, they did not try to treat these visual texts as a reflection of lectures that represent the real life academic situation.

The still photo texts exceeded the video texts in terms of the distracting reasons that test takers stated during the retrospective reports as Tables 4.12 and 4.14 showed above. Three distracting reasons were stated exclusively to still photo texts. Interestingly, the most frequent reason mentioned as distracting was also mentioned by some other test takers as helpful, which was the frequent pauses that are unique to the design of still photo texts. This design found to be strange and uncommon, and affected a group of the participants (37%) in a negative way, mostly because it was not moving normally like the case with video condition, and the several pictures that change at certain intervals forced some test takers to look at the screen each time they sense a change. It may be the case for those who were distracted by the pictures that the time they consumed while looking at the screen required them to increase the load of processing visual information and that might have led them to a decrease the processing of audio information. In this case, test takers would also potentially lose the opportunity to take any notes of what had been said by the speaker, therefore they end up with a situation where they lost some parts of the content of the lecture and they could not compensate for those parts from what they can see in the screen. This scenario seems likely because it also supports the idea of semantic congruity suggested by Suvorov (2013). Sometimes when the picture changes to the next one, it displays a new slide with different content from the previous one. This content might be related to a point that the speaker did not start to talk about yet at the time the test-taker inspected the screen, and started talking about it few seconds after they looked away back to their papers in order to take notes again. This condition happened frequently, and although caution was sought during designing the still photo texts, but because the photos needed to be changing systematically along the texts, this few seconds' difference between the display of a new slide and its related content was inevitable. Therefore, during the short time the test takers were inspecting the new picture, they found that what the slides display is not congruent with what the speaker says. This type of semantic incongruence between the visual and audio information can most likely be a source of distraction as suggested by Hu and Jiang's (2011) and Suvorov (2013).

Some test takers were also distracted by the still photo texts because the speaker in this condition does not move, so they couldn't get any type of assistance or clue from the speaker that can support their comprehension of the spoken text. This case is apparently falling in the category of static visual according to the taxonomy of visuals (Suvorov, 2013), which does not give any additional information to the viewer. In research of L2 listening tests with visuals, a still photo text that includes speakers is usually used with conditions like dialogues or conversations in order to set the scene of the communicative event which normally involves more than one speaker (Ginther, 2002). In the case of the present study, it was decided to include the speaker in the still photo texts in order to be systematized with the video texts and present a whole picture of the context of the lecture. As noticed earlier (Tables 4.9 and 4.10) that the speaker's appearance in the still photo texts was not helpful in any way as it was not mentioned by any of the test takers in the helpful aspects and reasons of the still photo texts. In addition, the presence of the speaker in still photo texts proved to be distracting to 33% of the participants, which gives an implication that the use of still photos to present a lecture or any type of monologue text might be better designed without including the speaker's image and presenting only the slides or any other supporting materials that can be normally used in that type of monologue. It seems that in this study the appearance of the speaker in the still photo texts falls in the category of rhetorical ineffectiveness within the spectrum of visual taxonomy, and that may explain the reason test takers perceived it as distracting. More precisely, the dimension of rhetorical effectiveness indicates that the visual should be presenting information in a persuasive manner. Therefore, when this type of visual information does not satisfy this condition, it can be regarded as a rhetorically neutral or in more intense cases, where the visual presented is found to be distracting -as some test takers reported-, it can be regarded as rhetorically sineffective. According to Kostelnick and Roberts (2011), the degree of rhetorical effectiveness of a visual is responsible for influencing the attention and interest of the viewers and triggering their background knowledge. As a result, that can affect the way those viewers perceive the visual. As the image of the speaker failed to evoke these qualities with the still photo condition, it seems possible to suggest that rhetorically ineffective visuals can be distracting to listening comprehension test takers.

On the other hand, the degree of effectiveness might be increased with some alterations to the design of the still photo condition. One suggestion that can potentially be beneficial is to present the speaker in the first photo and then presenting only the supporting materials (e. g., slides) in the following photos. The purpose is to set the scene at first, and then remove the source of distraction during the rest of the text. This suggestion needs to be investigated further in order to find out its effects on test takers' performance and perceptions.

Another distracting reason of still photo texts was found to be the disconnection between the audio and the visual input. This distracting reason cannot be separate from the previous two reasons (Frequent pauses and Speaker is not moving) stated solely for the still photo texts and again a continuous cause and effect relationship can be visualized. Hence it can be said that because the speaker is not moving, it is hard to connect the audio with the visual input. In addition, one can say it is hard to connect the audio with the video input because of the frequent pauses. This circular relationship generates similar interpretations to the distracting reasons found solely in the still photo texts.

While these results gave a clear picture of how the test takers perceived the video texts and the still photo texts during processing these texts, it is also important to investigate how these types of visuals contributed to the process of answering the listening test questions.

L2 test takers perceptions of visual information when answering the test questions.

The overall purpose of this part of research question 2 was to examine how did the test takers in this study use the visual information from the video and still photo texts when answering the test questions. In order to achieve this goal, the researcher asked each participant about the usefulness of visuals in answering each individual question in the video and still photo texts, and whether the visuals assisted them to answer the questions. Test takers were given the chance to elaborate in their answers but were asked eventually to give an explicit decision (Yes or No) of whether the visuals helped them to answer each question in order to correlate them with their scores. After collecting all data, a score of 1 was given to each question that was reported to be answered with the help of visuals, and a score of 0 was assigned to each question that was reported to be answered without the help of visuals.

The scores formed two groups: group A contains the scores for the test takers' perceptions of visual materials for answering the questions of the video texts, and group B contains the scores for their perceptions of visual materials for answering questions of the Still photo texts. For each individual test taker, the full score of perceptions for each type of texts is out of 10 (total number of questions in 2 video texts or 2 still photo texts). Using paired samples t-test, the test takers' perceptions of the usefulness of visuals in the two types of texts were compared. The results are presented below in Table 4.15.

Text type	M	SD	t	df	p
Video texts	4.33	2.52	1.74	29	.06
Still photo texts	3.60	1.84			

Table 4.15: Comparing the test takers' perceptions on the usefulness of visual information in answering the test questions in the video and still photo texts.

As displayed in the table above, test takers perceived visual materials in video texts (*M* = 4.33, *SD* = 2.52) to be more helpful in answering the test questions than the visual materials in still photo texts (*M* = 3.60. *SD* = 1.84). The difference in their perceptions was marginally statistically significant, *t* (29) = 1.74, *p* = .06, indicating that the magnitude of the difference between the means of the scores of perceiving visuals as helpful to answer the test questions in the two text types was relatively small.

This result shows that the test takers used the visuals in the video and the still photo texts to a different extent in answering the test questions, which in fact complies with the previous results of the cued retrospective data. It can be the case that the test takers' perceptions about the video and still photo texts were based on the way they answered the test questions.

However, it should be noticed that the difference in test takers' perceptions of the usefulness of visuals in answering test questions of the video and still photo texts was only marginal (*p* = .06), but that might be a result of the small size of the sample or even the limited number of test items. Therefore, it was decided to ask the test takers at the end of each interview to rank their preference of the three text types- video, still photos, and audio- that were used in the test (from the most preferred to the least preferred type) in order to get more general view of their perceptions. This ranking is a general one and based on the overall perceptions of the test takers. The results revealed large differences between the three conditions as Table 4.16 below shows. More than half of the test takers (53%) ranked the video texts as their most preferred type. They perceived it to be the most helpful regardless of the way this type of visual contributed to their answers of the test questions. In the second place came the still photo texts, with 27% of the test takers viewing it to be the most helpful, and lastly, only 20% of the test takers preferred the audio-only texts over the other two types.

TEST MODE	% OF PERCEIVING EACH TYPE AS MOST HELPFUL
VIDEO	53% (16 p)
STILL PHOTOS	27% (8 p)
AUDIO-ONLY	20% (6 p)

Table 4.16: Test takers reported experience of preference with the three text types.

These overall perceptions go in line with the previous results of the cued retrospective reports and how the test takers perceived the visual texts in processing and answering the test questions. Next, the sub-question of RQ 2 is answered.

RQ 2.1: **Is there any correlation between the test takers' perceptions and their scores?**

As test takers were asked during the cued retrospective report how useful the visual condition in each text was in answering the test questions, it was also decided to compare these perceptions with their actual test scores to find out if there was any correlation between them. In other words, the correlation would be informative to examine whether test takers who scored higher in a specific visual condition did actually perceive that condition as helpful.

To answer this question, two variables were used: test item scores on video and still photo texts for each participant and their individual perception scores for these items.

SPEARMAN'S RHO	PERCEPTIONS	
	r	*p*
VIDEO TEXTS	.45	.01
STILL PHOTO TEXTS	.22	.24

Table 4.17: Correlation between scores of the video and still photo tests and the test takers' perceptions of the helpfulness of visuals in answering their questions.

Spearman rank-order correlation was conducted in order to determine if there was any relationship between the test takers' scores in the video and the still photo texts and their perceptions of the helpfulness of visual materials in answering the test items. A twotailed test of significance indicated that there was a significant moderate positive relationship between the scores of the video texts and the test takers' perceptions of the visual materials being helpful to answer the test items (r_s (30) = .45, p = .01). This result suggests that when test takers perceive the visuals to be helpful in answering the test items, they tend to answer them correctly. On the other hand, the correlation between the still photo texts scores and the perceptions of the helpfulness of the visuals in these texts was also positive but weaker and statistically insignificant (r_s (30) = .22, p = .24). This indicates that test takers perceptions of the usefulness of still photo texts were not, most of the time, strongly correspondent to their answers of the test items. The two Figures below (4.4 and
4.5) present these relationships in scatter graphs.

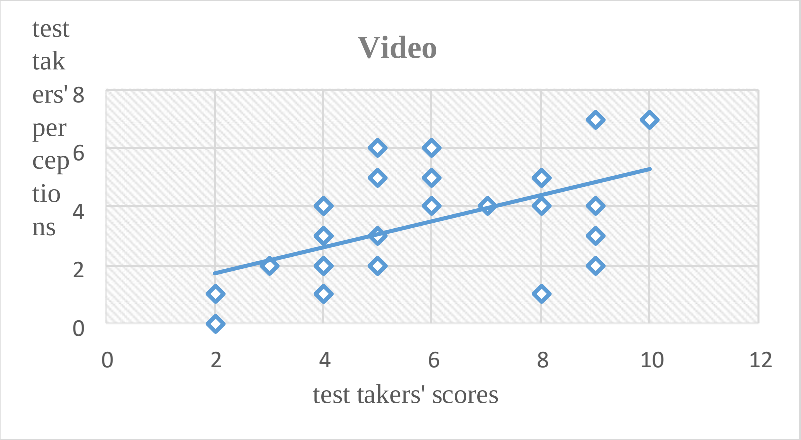


Figure 4.4: Correlation between test takers' scores and their perceptions of the video texts.

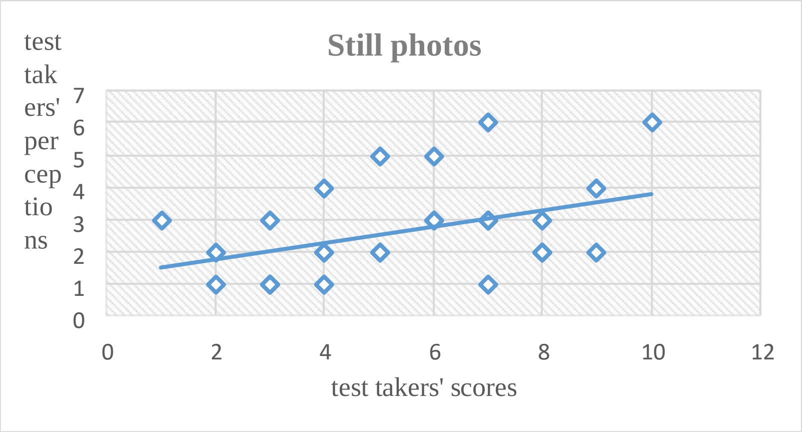


Figure 4.5: Correlation between test takers' scores and their perceptions of the still photo texts.

It can be said that the results of the Spearman rho correlation between the test takers' perceptions and their scores in the two visual text types are fairly consistent with the previous results. One possible interpretation of this result is that the test takers were more confident about their judgments on the usefulness of the video condition than they were with the still photo condition because it reflected the real-life situation of an academic lectures more than the still photo texts did, and that reinforces the idea that video texts present visual information that are semantically richer than the visual information in the still photo texts. Also, as the results of the test takers perceptions of the two visual conditions revealed that still photo texts were generally more distracting than video texts, then it might be the case that this distraction has also affected the test takers' judgements on whether they answered the test questions correctly or incorrectly because of the effect the still photo condition had on them.

Summary of research question 2:

To answer the first part of research question 2, the researcher coded and analysed verbal data collected from 30 test takers via a cued retrospective report after finishing L2 listening comprehension test. The results of the qualitative analysis showed that the test takers found helpful and distracting aspects in both video and still photo texts, and also specified reasons for finding video and still photos either helpful or distracting. However, the magnitude of their perceptions was different in the two visual conditions. The video texts surpassed the still photo texts in both the helpful aspects and the helpful reasons as stated by the test takers. In addition, their focus was in two categories: slides and speaker.

In other words, the test takers found the appearance of the speaker and the presence of the slides both to be helpful in the video texts, unlike with the still photo texts where the stated helpful aspects and categories by the test takers where only related to the presence of the slides or to the visual design of the still photo condition in general, without stating any factors related to the appearance of the speaker to be helpful. On the other hand, the still photo texts surpassed the video texts in the distracting reasons as stated by the test takers, in addition, distracting aspects were more commonly expressed with the still photo texts. Some categories were found to be distracting with both video and still photo texts, namely- both visual conditions were deemed by some test takers to over attract their attention to the screen leading them to miss parts from the spoken input. Also, a number of test takers found both visuals to be a new type of testing listening, that is different from what they commonly have (audio-only), and that resulted in distracting and confusing their test taking techniques which they were trained to use.

In addition, test takers were asked to evaluate how helpful the visuals were in answering the test items in the visual texts. The results revealed a marginally significant difference between the video and still photo texts with higher perceptions to the video texts as more helpful in answering the questions. Finally, those perceptions were correlated with the actual scores that test takers gained in the two visual conditions, using Spearman rho correlation coefficient test. The results revealed that the correlation was positive with both visual condition, but it was not statistically significant with the still photo texts revealing that the test takers' judgements on the usefulness of the still photo texts in answering the test items were not generally accurate.

Over all, the test takers had also the opportunity to rank the three listening conditions according to their preference. The video condition was the most preferred one among the three text types, with more than half of the test takers (53%) considered it to be the most helpful, followed by the still photo texts with (27%), and finally comes the audio condition with only (20%) considering it to be the most helpful.

Research question 3:

What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualised texts?

The third research question examined the viewing patterns of test takers (*n* = 30) in all visual trials, and compared these patterns in the video trails with the still photo ones. This question basically comprises three sections: first, the viewing patterns are examined via eye tracking measures. Second, these patterns are compared with the test takers' scores in the video and still photo texts, which were presented in research question 1, in order to investigate the relationship between them. Third, the viewing patterns are also compared against the test takers' perceptions of the visual texts, which were presented in research question 2.

To address this question, three basic eye tracker measures were used, these are fixation count, dwell rate, and total dwell time. The choice of these measures is believed to give a comprehensive picture of the way test takers interacted with the two types of visual texts and how important they found each one (Holmqvist et al., 2011). It is believed that the different types of visuals- video and still photos- might produce different cognitive processes (Field, 2012). This can be better explored by linking the eye tracking measures with other types of data like the qualitative verbal reports.

Before extracting the eye tracker data from the data viewer software, the researcher had first to initiate interest periods for the video and still photo texts in order to be able to apply the different eye tracker measures. The interest periods started from the beginning of the display of the video or still photo texts and lasted until the end of these visual texts. It excluded the display of the test items -before and after the texts- as the focus is only on the viewing behavior during the visual texts display, and not while previewing or answering the test items.

Beside the interest periods, interest areas (AOI) needed to be established for each visual text. Initially, one area of interest was created for each visual text in order to calculate the three eye tracking measures within that area. The AOI formed border lines that framed the whole scene of the visual text (Figure 4.6). With the purpose of obtaining a deeper level from the data, the first measure -fixation count- was used in two ways: first with one AOI for each visual text as explained earlier, in order to compare the viewing behavior between the video and still photo texts as shown in Figure 4.6. Second, two AOIs were created within every visual text, with the first capturing the speaker's zone only and the second capturing the slides' zone as shown in Figure 4.7. This technique helps to measure the way test takers focused in these two major visual zones, and consequently it can give an indication about how important each zone was to the test takers, and whether they placed more attention on one area more than the other comparing the video and still photo texts.

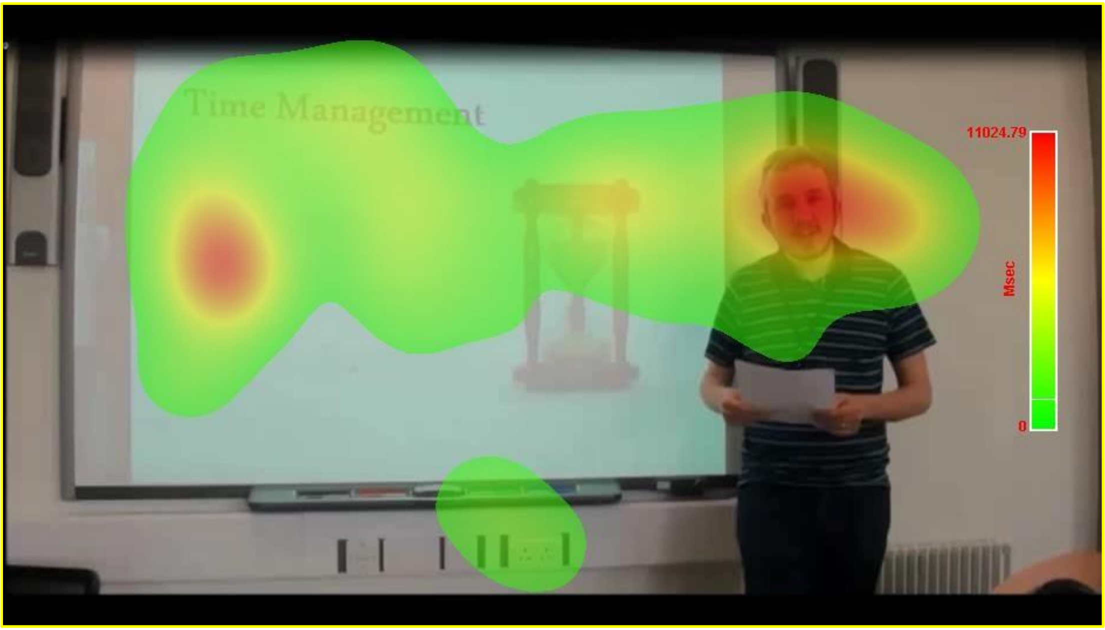


Figure 4.6: The yellow borders show the AOI in one of the visual texts.

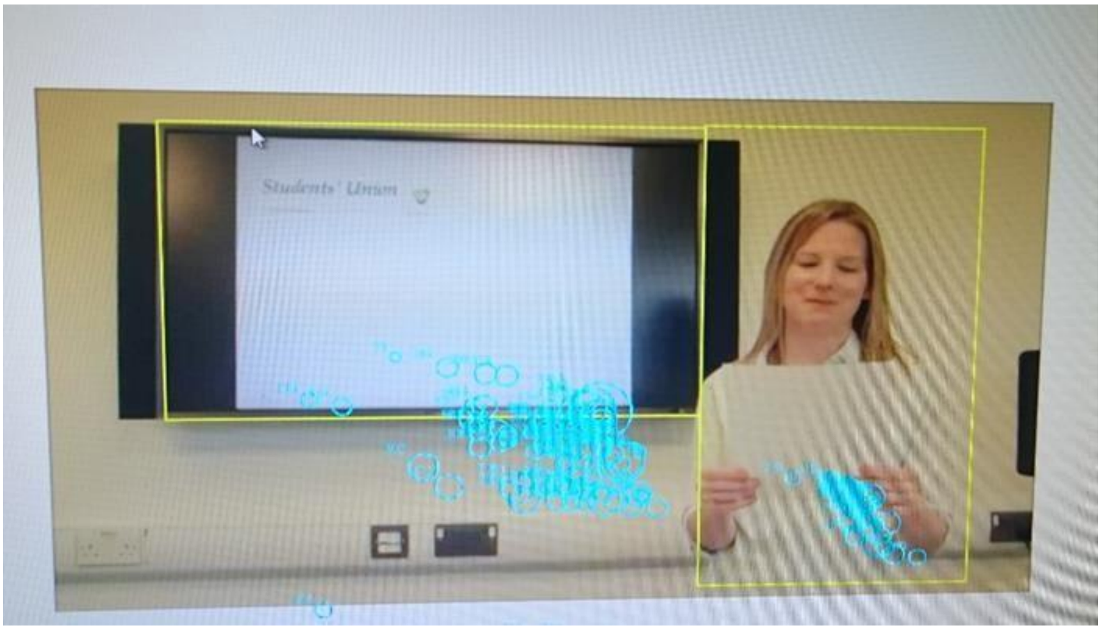


Figure 4.7: The two AOIs framing the speaker's zone and the slides' zone in the visual texts.

It is important at this stage to justify the use of the fixation count in this study and not fixation rate which is more commonly used in researh. According to Holmqvist et al., (2011) fixation counts can be used to compare fixations of texts that are of equal size in terms of time and magnitude. If this condition was not met, then fixation rate should be used instead. This condition obviously would be the reasonable choice if the texts had been looked into individually, since the six texts used in the test are of different lengths. However, the use of fixation count was considered instead of fixation rate because the goal of the third research question was to investigate whether there is any difference in test takers' watching behavior in the two visual types, namely video texts and still photo texts. Technically, the two visual types use exactly the same texts, with equal amount of watching, as explained in Figure 4.8 below, which means they are of equal size and length in total.

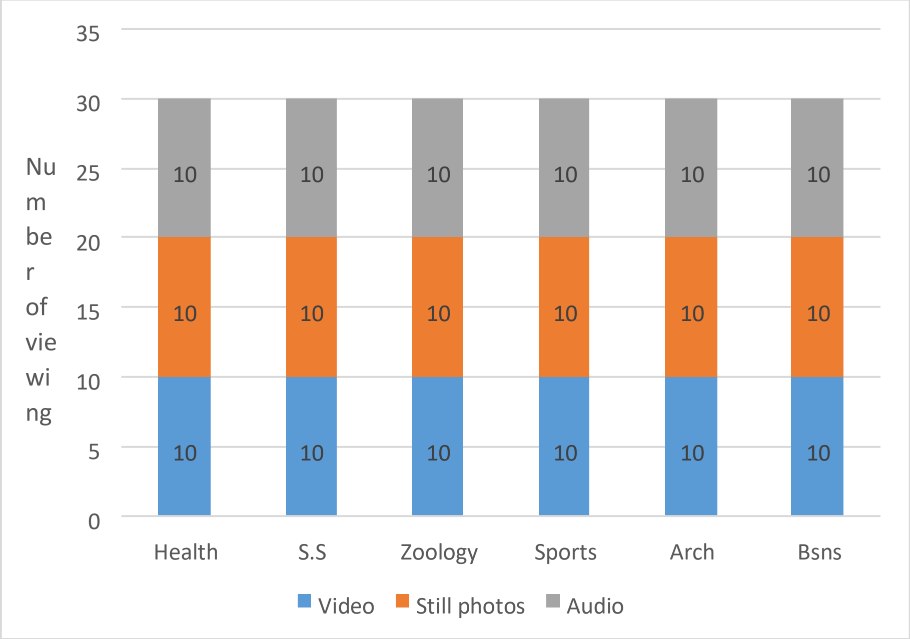


Figure 4.8: Number of viewings for each of the three text types.

Some descriptive statistics were calculated before answering research question 3. These show how the three eye tracking measures mentioned previously were associated with each of the 6 listening texts, regardless of being in video or still photo condition.

Fixation count AOI

	M	SD	Skewness	Kurtosis
Health	199.60	139.5	1.160	.572
Social Science	329.45	249.9	.111	-1.591
Zoology	380.15	179.3	-.119	-.345
Sports	257.95	179.3	.881	.123
Architecture	241.75	161.9	.345	-1.173
Business	243.95	141.9	-.016	-1.455
Average	275.47	175.3	0.393	-0.644

Table 4.18: Descriptive statistics for fixation count with the 6 visual texts.

Dwell rate m^{-1}

	M	SD	Skewness	Kurtosis
Health	22.14	16.14	1.235	.836
Social Science	25.15	22.16	.713	-.857
Zoology	20.24	20.23	.935	-.384
Sports	20.17	19.16	1.206	1.327

Architecture	23.19	22.14	.453	-1.167
Business	28.22	26.21	.375	-1.519
Average	23.19	20.84	.986	.372

Table 4.19: Descriptive statistics for dwell rate with the 6 visual texts.

	Total dwell time %			
	M	SD	Skewness	Kurtosis
Health	55.27	17.54	1.724	3.920
Social Science	49.60	25.00	1.036	-.213
Zoology	64.27	23.40	1.121	.495
Sports	59.83	20.45	1.342	.912
Architecture	60.45	21.78	1.472	1.399
Business	53.48	21.19	1.092	1.054
Average	57.15	21.89	1.297	1.261

Table 4.20: Descriptive statistics for total dwell time with the 6 visual texts.

The descriptive statistics presented above show that test takers (*n* = 30) fixated their eyes on the screen at an average fixation count of 275.47 per AOI. They also re-visited the screen (Dwell rate) at the rate of 23.19 per minute and spent on average 57.15% of the total time watching the visual texts. However, it worth mentioning that the values of the standard deviation with the three measures were relatively high referring to the large individual differences among test takers regarding their watching behavior. Skewness and kurtosis values were all less than 2, which indicates that the data of the eye tracking were normally distributed.

The first part of research question 3 was addressed by comparing the viewing patterns of the test takers in both the video and still photo texts. Paired samples t-tests were run with the three main eye tracking measures (i.e., fixation count, dwell rate, and total dwell time). The first measure- fixation count- was analyzed further according to the two AOIs in each visual text. As explained earlier, it is considered important to investigate the watching behavior patterns of test takers, not only in the two types of visual texts (video and still photo), but also to go deeper and inspect how they fixated their eyes in the first and second areas of interest, where the first AOI includes the speaker’s zone (AOI 1) and the second includes the slides’ zone (AOI 2). The results of the t-tests are shown in Table 4.21 below.

Measure	visual	M	SD	t	p	df	Effect type size η2
Fixation count	V	315.65	167.97	2.50	.01	59	.32
	SP	253.50	172.16				
Fixation count AOI 1	V	155.40	101.54	.21	.82	119	.03
AOI 2	V	157.88	85.45				
Fixation count AOI 1	SP	105.26	91.62	2.29	.02	119	.41
AOI 2	SP	125.52	99.32				
Dwell rate	V	24.10	19.23	.56	.58	59	.07
	SP	22.04	17.07				
Total dwell time	V	57.43	20.12	4.13	.00	59	.53
	SP	49.77	22.69				

Table 4.21: t-test results for the three eye tracking measures used in the study.

Primarily, it is clear from the results of the *p*-value above that test takers performed similarly with some eye tracking measures (statistically insignificant difference), while there were some statistical significant differences in other measures. Looking first at the results of the second eye tracking measure which showed that the dwell rate of the video texts (*M* = 24.10, *SD* = 19.23), was higher than the dwell rate of the still photo texts (*M* = 22.04, *SD* = 17.07), but the difference was not statistically significant, *t* (59) = .56, *p* = .58. partial η^2 = .07, indicating that only 7% of the variation in the dwell rate could be attributed to the visual type.

On the other hand, two of the eye tracker measures revealed statistically significant differences between the video and still photo texts. First, results of fixation count showed more fixations in the video texts (*M* = 315.65, *SD* = 167.97) than in still photo texts (*M* = 253.50, *SD* = 172.16) with a *t*-test result of *t* (59) = 2.50, *p* = .01. partial effect size η^2 = .32

indicating that 32% of the variations in the fixation count could be attributed to the visual type. Furthermore, the fixation count measure was also analyzed according to the two AOIs in each video and still photo text by conducting paired samples *t*-tests within each visual type in order to see how the listeners viewed the speaker's zone and the slides' zone. Some differences were revealed from this analysis. With the video texts, fixation counts within the first AOI which represents the speaker's zone (*M* = 155.40, *SD* = 101.54) was very close to the number of fixations within the second AOI that represents the slides' zone (*M* = 157.88, *SD* = 85.45), with no statistically significant difference at *t* (119) = .21, *p* = .82, and partial effect size of $\eta^{2\,=}$ .03, indicating that the AOI is responsible for only 3% of the variations in the fixation count. On the other hand, with the still photo texts the results of fixation count within the two AOIs were different. The *t*-test results revealed that the speaker's zone AOI (*M* = 105.26, *SD* = 91.62) had statistically significant lower amount of fixations than the second AOI which represents the slides' zone (*M* = 125.52, *SD* = 99.32), with a *t*-test result of *t* (119) = 2.29, *p* = .02, and partial effect size of η^2 = .41, indicating that 41% of the variation in the fixation count can be attributed to AOI.

The analysis of the last measure of eye tracking revealed that a statistically significant difference was found between the total dwell time of the video texts (*M* = 57.43, *SD* = 20.12) and the total dwell time of the still photo texts (*M* = 49.77, *SD* = 22.69) with a result of *t*-test of *t* (59) = 4.13, *p* = .00, and partial effect size of η^2 = .53, indicating larger effect size from the previous three measures, which suggested that 53% of the variation in the total dwell time can be attributed to the visual type.

Zooming in on these results reveals that the test takers in general treated the two different visual types differently. The fixation counts and total dwell time results show the magnitude of this difference with higher values in both measurements related to the video texts over still photo texts. The results of the eye tracking data showed clearly as demonstrated above with the total dwell time results (Table 4.21) that the test takers in this study spent significantly longer time watching video texts (57% of the total time of videos) than still photo texts (50%). One can attribute this difference in viewing behavior regarding the total amount of time spent watching to the possibility that test takers found the video texts more interesting and engaging than still photo texts. This interpretation in fact reflects what Holmqvist and his colleagues (2011) believe to be an explanation of the total dwell time, which can indicate that viewers normally spend longer time watching visuals if they find them highly informative and interesting. However, without correlating these results with the test takers' perceptions, this interpretation cannot be agreed on. For instance, it can also be logically suggested that spending longer time watching the visual might refer to difficulties in interpreting some visual aspects like unclear textual information in the slides or even a strange behavior from the speaker. Also, it might be the case with some test takers that they were actually focusing their foveal on the screen (i. e., overt attention) while the text was playing but their actual attention (i. e., covert attention) was not on the aspect they were looking at. This possible interpretation in fact violates the eye-mind hypothesis suggested by Just and Carpenter (1984) which states that what a subject is looking at has a strong correlation to what he/she is thinking about. This issue is investigated more in the last part of research question 3 where a correlation test is conducted to find out the relationship between the eye tracking measures and the test takers' perceptions and their verbalization of the way they watched the visual texts.

This result of higher amount of the total dwell time with the video texts over still photo texts appears to be similar to the results in Ockey's study (2007). Ockey however did not use these eye tracking measure as no eye tracker device was involved in his study. The total watching time of his participants was measured using a stop watch while observing a video record of each participant's watching behavior. No percentage of the total watching time was stated in that study, nevertheless Ockey stated in general that "test takers have little engagement with still images while on the other hand, most do engage with the video stimulus" (2007, p. 527).

Similarly, the fixation count result also indicated that in total, the test takers fixated their eyes on the video texts significantly more than the still photo texts as revealed by the difference in the means of that measure (i.e., 216.10 for video texts, and 54.12 for still photo texts). Fixation count, as specified by Holmqvist et al., (2011), can be a sign of finding the target AOI semantically informative, and the number of fixations found to be increasing with objects that viewers regarded as important (Jacob and Karn, 2003). Hence the results of fixation count data might be an indication that test takers found the video texts highly more semantically informative than still photo texts as suggested by the large difference in the *t* value. However, unless these results are correlated with other type of analysis (e. g., cued retrospective reports data), high fixation count may also refer to difficulty in processing information in the visual stimuli.

The results of fixation count per AOIs in video and still photo texts revealed another dimension to the way test takers watched these visuals. Besides the fact that they fixated their eyes on the video texts more than on the still photo texts, the analysis per AOIs showed the test takers did fixate their eyes on the slides' zone (AOI 2) more than the speaker's zone (AOI 1) in the still photo texts, while no such difference was found in the video texts, as they watched both zones to a similar extent. This result in fact is in harmony with the previous result of total fixation count per text, and it suggests that the lower amount of fixations on the still photo texts might be attributed to the fact that test takers focused largely in one zone (AOI 2: the slides) and mostly neglected the other (AOI 1: the speaker) in the still photo texts while placing more attention on both zones in the video texts resulting in longer watching time. It might be the case that with the still photo texts participants did not find the picture of the speaker who is not moving in congruence with the audio to be of any semantic importance and consequently did not support their comprehension. Therefore, they preferred putting more attention on the slides that possibly would offer more information which can be either textual or pictorial. On the other hand, when watching video texts, it is possible that test takers divided their attention between the two dominating zones (speaker and slides) because their visual content was dynamic and semantically congruent with the audio information. In fact, semantic congruence seems to play a significant role in listeners' attention and comprehension of visual spoken texts. As discussed earlier, results from other studies had also shown the importance of this dimension which resulted in more attention to the visual aspects that go in line with the audio information. In Suvorov's (2013) study about using content and context videos in listening comprehension tests, he found that participants who perceived content videos to be more helpful are those who focused more in the movement of the speakers in the screen and how they interact with the slides or any other type of support materials provided. They considered this type of videos to be semantically congruent and regarded the dynamism of the speaker as a scaffolding element that helps to link the visual and audio information together.

Regarding the dwell rate measure, a different result from the other two eye tracking measures appeared. The data indicated that test takers did not treat the video texts and the still photo texts differently with respect to the number of re-visits to these two visual types. Researchers like Jacob and Karn (2003) and Holmqvist et al., (2011) agreed that the dwell rate measure commonly reflects the importance of a specific AOI to the task that accompany it. Hence, according to this indication, the dwell rate results imply that test takers found the video and still photo texts both important for completing the task. In fact, there could also be different interpretations for this result, bearing in mind that the two different visual types employed exactly the same texts and tasks. It might be the case that the revisits were affected by the type of tasks with each text rather than of being in a video or still photo mode, for instance, a task that required the test takers to fill-in gaps of a specific diagram like the food pyramid in the Health text, or the table of features and consequences according to specific time zone in Social Science text. If this was the case, then it can be argued that participant might have revisited specific texts (e. g., Health and Social Science texts) in both video and still photo mode more than other texts because of the impact of the task rather than the impact of visual type. Another interpretation which is based on the researchers' own observation while test takers were being tested, is that lots of the revisits happened while the test takers were taking notes so they had to look away from the screen and into the paper provided in order to record their notes and then look back into the screen. In this case, it can be said that the number of revisits might be affected by the amount of notes taken, and that would certainly differ from one test-taker to another. Finally, it is also possible that treating the two visual types similarly was due to the fact that there was only one AOI in each text and there were not any other areas that would attract the test takers' attention, leaving no other choice for the test takers but to look each time to the same AOI.

Looking at the eye tracking data shows other important results regarding the video and still photo texts. It is obvious from Table 4.21 above that there were large amounts of variation in watching behaviors among the test takers as specified by the differences in standard deviations in all the three measures of eye tracking. These differences suggest that participants behaved largely different regarding: 1] how many fixations on each video and still photo text (and also per each AOI in the video and still photo texts), 2] how many revisits they executed with each video and still photo text (per minute), and 3] how long they spent in total watching each video and still photo text. Beside the fact that large differences in the watching behaviors were found with both the video and still photo texts, it is interesting to notice that the values of standard deviation were particularly higher with the still photo texts in both fixations count and total dwell time, which indicate that the data were further spread in these two measures with still photo texts more than with video texts. For examples, the data from the fixation count revealed that participant 9 had a very low number of fixation with one of the still photo texts (i. e., Business, 38.00), while participant 18 on the other hand fixated his eye for a largely bigger number on the same still photo text (731.02). Similarly, participants 1, 13, and 27 spent around 75% of the total time watching one of the still photo texts, while participants 5, 20, 24, spent only about 21% of the time watching still photo texts. This variation in the test takers' watching behavior seems to go in line with the findings of previous studies in the field of using visuals in listening comprehension tests. Most importantly the research done by Ockey (2007), Wagner (2006b, 2010), and Suvorov (2013) who also found considerable variations in the rates of watching behavior of their subjects. However, it is important to mention that among those three studies only Suvorov's (2013) study employed eye tracker device in recording the participants' watching patterns, where the other two studies gathered the watching behavior data by simply videotaping the participants while they were taking the test and then measuring the amount of looking at the screen for each test-taker individually. It is also important to highlight the fact that only in Ockey's (2007), the types of visuals used were video and still photo texts, while the other two studies only investigated the effects of different types of video texts. Thus, by using eye tracking technology to compare test takers' viewing behavior of two different types of visual materials, video and still photos, the current study makes new and important contribution to the field of L2 listening assessment.

Regarding the reasons of variations in the test takers' watching patterns, there can be several motives that might have led to such big variation. One interpretation could be related to the quality of the eye calibration process done at the beginning of each test. This process differed to some extent in its execution among the test takers. While with some of the subjects the calibration was performed in a very smooth way and the camera was able to capture and record their pupil immediately, a number of other participants struggled with the process of calibration, and it needed to be repeated several times before the camera could spot the subject's pupil. Commonly, the quality of the calibration process can be affected by some physical characteristics of the subjects. For instance, some of the factors that can disturb the calibration quality are the very dark eye pupil, like the case with some of the Arabic test takers, or the very narrow eyes with droopy eyelids like the case with some of the Asian test takers, where their eyelids conceal part of their pupil. There was also the problem of the short-sighted participants who were wearing thick glasses, which made it hard for the process of calibration to be completed seamlessly, and so it very often needed to be repeated. However, despite the existence of these problems, their extent was acceptable and did not prevent, in most cases, the subjects from proceeding to the test. In other words, the cases that were seriously threatening the data quality had been excluded before taking the test as the calibration would not be successful even if performed repeatedly.

Beside those physical characteristics which are regarded as a trivial problem to the calibration quality, the large variation in viewing patterns might have also be affected by the test takers' individual differences which could affect the way they interacted with the different visuals. In order to investigate whether the individual differences and other factors contributed to the variation of the results, it is important to associate these eye tracking results with the other results revealed earlier, namely the test takers' test scores and their perceptions towards the two visual types in order to find out whether there was any relationship between them.

Correlation between the eye tracking measures and the listening test scores.

The second part of research question three investigates whether any relationship can be detected between the three eye tracking measures discussed above and the L2 listening test scores gained by the test takers. In order to address this query, Pearson correlation coefficient (*r*) was used to assess the degree of relationship between each one of the three eye tracking measures and the test takers' final scores on the video and the still photo texts. Table 4.22 below presents the correlation results.

EYE TRACKING MEASURE	VIDEO SCORES		STILL PHOTO SCORES	
	r	*p*	*r*	*p*

FIXATION COUNT	-.08	.65	-.01	.93
DWELL RATE	.15	.20	.43	.19
TOTAL DWELL TIME	.22	.07	.09	.28

Table 4.22: Correlation between the three eye tracking measures and the test scores in video and still photo texts.

As revealed by the above results, there was no relationship between fixation count and the scores of video texts (*r* = -.08) neither between fixation count and the scores of still photo texts (*r* = -.01), since the values of *r* were closer to zero. Regarding dwell rate, the correlation between this measure and the video score (*r* = .15) indicated a very weak relationship, while a moderate relationship was found between dwell rates and still photo scores (*r* = .43), however this relationship was not statistically significant at *p* = .19. Finally, with total dwell time the Pearson correlation coefficient found a weak relationship with the scores of the video text (*r* = .22), and no relation was found between this eye tracking measure and the still photo texts scores as the correlation value (*r* = .09) was close to zero.

Having a look at these results reveals that the test takers' viewing patterns did not have any relation with their scores. That is to say, the number of fixations the test takers proved to have on the screen, the number of revisits to the AOIs, and the total amount of time they consumed watching the visuals were not actually the factors that affected their scores as no statistically significant relationship was found in the results. This implies that there was a variation in the viewing patterns of the video and still photo texts among most of the test takers- namely, those who achieved higher scores as well as those who received lower scores in both visual types. Hence, it seems that the nature of interacting with the video and still photo texts did not have an effect on the scores of test takers.

As it seems that the test scores were not affected by the presence of visuals in the video and still photo texts, it is important nonetheless to investigate whether these patterns produced by the viewers had any connection with the way the they processed the texts. In other words, the third part of research question three attempts to examine the correlation between the three eye tracking measures and the test takers' perception of the two visual texts.

Correlation between eye tracking measures and test takers' perceptions of video and still photo texts.

Similar to the procedure followed in the previous part of research question three, a Pearson correlation coefficient (*r*) was conducted in order to find out if there was a relationship between the test takers' viewing patterns in the video and still photo texts and their perceptions about the two visual texts. The results of this analysis are presented in Table 4.23 below.

EYE TRACKING MEASURES	PERCEPTIONS OF VIDEO TEXTS		PERCEPTIONS OF STILL PHOTO TEXTS	
	r	*p*	*r*	*p*
FIXATION COUNT	.12	.52	.04	.83
DWELL RATE	.22	.18	.76	.06
TOTAL DWELL TIME	.83	.04	.16	.09

Table 0.23: Correlation between viewing patterns and test takers' perceptions of the video and still photo texts.

The correlation analysis results showed a very weak relationship between fixation count and test takers' perceptions about the usefulness of the video texts (*r* = .12) and no relationship emerged with their perceptions about the still photo texts as the result is closer to zero (*r* = .04). Regarding the dwell rate, another weak relationship was revealed between the test takers' perceptions about how useful the video texts were and their dwell rate measures (*r* = .22). On the other hand, the correlation results of this eye tracking measure revealed a strong relationship with how the test takers perceived the still photo texts (*r* = .76), and this relationship was marginally significant at *p* = .06. Another strong relationship was found between the total dwell time measure and perceiving video texts as useful (*r* = .83), and this relationship was statistically significant at *p* = .04. The last correlation which is between the test takers' perceptions of still photo texts and their total time of watching was very weak (*r* = .16) indicating the absence of any connections between their total time of watching and their perceptions of the usefulness of this visual text.

These findings revealed that in general, the test takers' perceptions about the usefulness of the two visual types were not related to their viewing patterns with the exception of two patterns-namely, the dwell rate with the still photo texts and the total dwell time with the video texts. The strong correlations in these positions however should be interpreted with caution, bearing in mind that correlation between them does not necessarily mean that one variable was a cause of the other. Considering the marginally significant value of the correlation between the number of revisits to the still photo texts and the test takers' perceptions of these texts as useful, one can suggest that this type of visual texts might have served an organisational purpose that affected the viewing behaviour of the participants. In other words, it might be the case that the nature of displaying the still photos at regular intervals triggered the viewers to systematically control their viewing behaviour and look at the screen every time they sense a change in the screen. This way had possibly given those test takers who perceived the still photo texts positively a feeling of control over the visual, which might have helped in minimising the feeling of distraction. If this was the case, then it might be said that the relationship was a causational one since the nature of the visual caused the test takers to view it in a specific manner. That is, to look at the screen whenever the photo changes. On the other hand, one can also argue that there would always be a possibility that this strong correlation was obtained by chance, and that is because of the relatively small size of the population in the study. It might also be the case that a different factor caused the raise in the number of revisits to the screen with those who viewed the still photo texts as useful, like for instance the need for taking more notes. Undoubtedly, some test takers did take notes more than others, and it seems to be likely that the number of revisits to the screen while the still photo texts were playing had coincided with the changing pictures because that helped the test takers to summarise the new information whether it was textual or pictorial, presented by the new photo so they can write more notes possibly with higher quality without the distraction of the moving speaker and that may possibly led them to positively perceive the still photo texts. However, to know whether this was the case, a closer examination of the test takers' notes taking behaviour would be necessary.

Similarly, the strong correlation found between the total dwell time and the test takers' perception of video texts as useful should also be interpreted with caution. It is possible that test takers who looked for a greater amount of time into the video texts screen had a greater chance to benefit from the visual and non-verbal information available either in the slides or in the speakers' body language, and for this reason they ended up seeing this type of visual as useful and supporting to their comprehension. It would also be rational to suggest a reversed interpretation, that is test takers who initially felt more comfortable watching video texts than still photo texts decided spending longer time watching the videos. These results go in line with the findings from Wagner's (2006b) study, who found that the participants who watched the video texts for longer time than others who watched them for shorter amount of time did eventually find the videos useful and more helpful in interpreting the aural content of the visual text. However, it should be mentioned that no eye tracker device was used in that study and the total amount of watching time was calculated using a video camera that recorded the participants' orientation to a large screen.

Summary of research question 3:

This last question of the study investigated the test takers' viewing behaviour using the eye tracking technology. It sought to draw patterns of test takers' viewings to the video texts and still photo texts using three main eye tracking measures- namely fixation count, dwell rate and total dwell time. These eye tracking measures were compared to the two types of visual texts using paired samples *t*-tests. In addition, the fixation count measure was also calculated according to the different AOIs in each video and still photo text, which consisted of two AOIs per text, the first represents the speaker's zone and the second represents the slide's zone. The results revealed that test takers fixated their eyes on the video texts more than the still photo texts and also spent a total amount of time watching the video texts more than the still photo texts, and in both cases the difference was statistically significant (*p* < .05). This seems to present a good evidence that test takers found video texts more informative and possibly more interesting than still photo texts. In the meantime, the rates of revisits to the screens of both video and still photo texts did not prove to be different. With regard to the fixation count of the two AOIs per visual text, the results showed that test takers fixated their eyes on both AOIs in video texts to a similar extent, while with the still photo texts they tended to fixate their eyes on the slides' zone more than the speaker's, and the difference was statistically significant at *p* = .02.

In addition to the *t*-tests, the eye tracking data was also correlated with the test takers' scores in both the video and still photo texts and likewise, correlated with their perceptions about these two visual types. The analysis of the results, using Pearson correlation coefficient, revealed that there was no statistically significant relationship between the test takers' scores in the video and still photo texts and their viewing patterns. This result suggests that test takers who scored highly and those who scored low had possibly treated the visuals similarly. Lastly, the correlation results of the test takers' viewing patterns and their perceptions of the video and still photo texts revealed that there was a relationship between the total amount of time that test takers spent watching video texts and their perceptions of finding these texts useful. Possibly because they found the video texts more interesting as suggested by Holmqvist et al., (2011) and less distracting than still photo texts in general. The results also showed that there was a marginally significant relationship between the rate of revisiting the still photo texts (dwell rate) and the test takers' perceptions about these texts as useful. Several possible interpretations can be linked to this relationship like for instance the probability that test takers who perceived still photo texts positively revisited the screen more because these texts presented a kind of summary to the whole situation so they felt comfortable looking at it every now and then to support their understanding and nourish their note taking process. Obviously, these results are not definite and more research is still needed. However, despite the limitations, using the eye tracker in this study does establish some practical and theoretical insights about the

use of visuals and test takers' behaviour in listening comprehension tests.

4.4. Summary of chapter four:

This chapter has investigated the three main research questions and presented the analysis of the collected data in an attempt to answer these questions. The questions explored three fundamental areas that altogether aimed at drawing a more comprehensible picture of the process of taking a L2 listening comprehension test accompanied with visuals.

These areas are basically categorised as: a) test takers' scores, b) test takers' perceptions, c) test takers' viewing patterns. Particularly, the analysis performed to answer research question one revealed that test takers' scores on the video texts were significantly higher than their scores on the audio texts, while there was not statistically significant difference between their scores on the video texts and the still photo texts, nor between their scores on the still photo texts and the audio texts. Next, the results of research question two showed that test takers perceived some useful as well as some distracting aspects in both video and still photo texts, and they also provided reasons for perceiving each visual type as useful or distracting. In addition, the results revealed that test takers perceived video texts to be, to a small extent, more helpful than still photo texts in answering the test questions, as the difference was marginally significant. Lastly, results from the third research question which dealt with the eye tracker data revealed that test takers had in general fixated their eyes on the video texts to a larger extent than they did with the still photo texts. Similarly, they spent longer time in total watching the video texts than the still photo texts. Analysis of the two

AOIs in each visual text (speaker's AOI, and slide's AOI) revealed that test takers tended to fixate their eyes more on the slide's zone than on the speaker's zone in the still photo texts, while placing similar number of fixations on both zones with the video texts.

Furthermore, the correlation results between the test takers' viewing patterns and their test scores in the two visual types showed no significant relationship between them. The final part of research question three sought the relationship between the test takers' viewing patterns and their perceptions about the video and still photo texts. The results suggested a marginally significant correlation between the number of revisits (dwell rates) to the screen of the still photo texts and their helpful perceptions of theses texts. There was also a significant relationship between the total amount of time (total dwell time) that the test takers spent watching the video texts and their perceptions about these texts as useful.

Chapter Five

Discussion

The study produced a number of useful insights into the way test takers performed in L2 academic listening comprehension test. The test employed two types of visual materials, namely video texts and still photo texts, in addition to the traditional audio only texts. The insights will be discussed according to each research question in turn.

5.1. Discussion of research question 1:

To what extent does the performance of L2 test takers differ on a listening test with a] video texts, b] still photo texts, and c] audio-only texts?

Data analysis of research question 1 using repeated measures ANOVA indicated that test takers performance was superior in video texts as compared to audio only texts with a statistically significant difference. Performance in still photo texts was also higher than performance in audio only texts, however the difference was not statistically significant. The results reflect the priority of both visual text types, especially video texts, compared to the audio only, which advocates the idea that visual materials can positively contribute to test takers' performance in L2 academic listening tests (Field, 2012; Ginther, 2002; Rost, 2013; Suvorov, 2013; Wagner, 2006b, 2010). As the combination of multiple channels of presentations (audio and visual) to deliver a spoken input resulted in improving the test takers' performance, it reflects a positive case of interaction between the components of the communicative language ability framework as presented by Bachman and Palmer (1996). More precisely, it is well maintained that the ability to process meaning from the different sources of information (i. e., audio and visual) comprises the ability to effectively employ strategic competence in order to integrate both audio and visual information with the linguistic and non-linguistic knowledge of the listener, as well as with the specific context of the speaking event (Vandergrift, 2007; Wagner, 2006).

The results also represent a case of support to the call for revising and expanding the

L2 listening construct as proposed by some researchers (Field, 2012; Ockey, 2007; Wagner, 2006a, 2006b, 2007). The unique characteristic of visual materials used in video and still photo texts (e. g., pictures, diagrams, speakers' body language in case of video texts) reflected essential elements of the target language use domain- which in this case academic lecture listening- that is regarded as a central quality of language tests (Bachman & Palmer, 1996). This reflection of TLU characteristics proved to improve test takers performance as indicated by the scores in the visualized texts. Because the L2 listening test used in this study is designed to predict test takers' language ability in academic domains, then the visual component of that domain should be included as part of the construct of L2 listening ability. Traditional audio only L2 listening tests have neglected the use of visual materials as a natural element of the spoken input, and this as a result could affect the validity of the inferences made by test designers on the language ability of test takers. Preventing L2 listeners from using their ability to process and use visual information during listening to spoken language would be perceived as a construct underrepresentation variance (Messick, 1996). If the purpose of L2 listening tests is to assess listener's' ability in a target language use domain that included visual materials, then excluding them in listening tests might be a serious threat to the validity of the L2 construct (Bachman and Palmer, 1996; Field, 2012).

The positive effect of including visual materials on test takers' performance as indicated by the results of this study corresponds to the growing body of authenticity in L2 listening literature. Namely, the use visual materials can increase the degree of authenticity of the test (Field, 2012; Suvorov, 2013; Wagner, 2006b), which is a central quality of language tests (Bachman and Palmer, 1996) as the degree of correspondence between the features of the target language use domain (i. e., academic lectures) and the test task is improved. In natural academic lecture listening situations, student would benefit from the supporting visual materials that lecturers normally use like Power Point slides or smart screens. They would also benefit from seeing the lecturer's kinesics (e. g., body language, gestures, face expressions) while listening to the lecture. Visual materials in L2 listening tests thus serve to better mirror this realistic situation (Wagner, 2006b). This result also corresponds to the believe that increasing the level of authenticity can possibly increase test takers' positive performance and perceptions of the test (Bachman and Palmer, 1996).

The case of comparing the effects of three types of listening texts (video, still photo, and audio only) on test takers performance has not received the deserved attention in L2 listening tests research. Previous studies on the effects of visual materials on test takers' performance focused mostly on the impact of video texts compared to audio only (Londe, 2009; Progosh, 1996; Suvorov, 2013; Wagner, 2006b, 2010). Investigating the role of still photo texts is very scarce despite its importance as being currently used in some standardized tests, like computer-based test TOEFL. Ockey (2007) compared listeners' engagement with video and still photo texts without focusing on their possible effect on test performance. Ginther (2002) conducted a study to compare the effects of still images on performance of L2 test takers, and found that still images can improve performance when the content of those images complement the audio input. However, no comparison was done with the video texts. Suvorov (2011) compared the effects of including single still image, video, and audio only on test takers' performance. In his study, Suvorov found that videos had a detrimental effect, while no significant effect of single image was found on test takers' performance. It is clear then how the impact of using multiple photos on the performance of test takers in L2 listening tests can hardly be compared with the available body of research.

Regarding test takers performance with the video texts, the study results echoes some of the previous research. Wagner (2006b, 2010) also found the use of video texts (through two types of stimuli: lecture and dialogue) can facilitate comprehension and thus performance. Sueyoshi and Hardison (2005) found the impact of videos with visual clues like face expressions to be positive and increased participants' scores. On the other hand, the increase in test takers scores in video texts as compared to audio only as revealed by the current study contradicts with the results of other studies. Suvorov's (2009) study revealed that video texts had a harmful effect on test takers' performance compared to audio only texts. Coniam (2002) did not find any significant difference between participants' scores in video and audio only texts. Similarly, Gruba (1993) found that students' performance on two versions of academic lecture listening texts - video and audio only- was not statistically different.

In sum, the results of research question 1 implies the need to reconsider the L2 listening construct that is used with most standardized L2 listening tests in order to include the visual component. The results suggest that the use of visual materials can improve text authenticity which consequently might increase test takers performance. This result relates to what Bachman and Palmer (1996) stressed that language tests should be designed to "elicit test takers' best performance" (p. 66).

5.2. Discussion of research question 2:

How do L2 test takers perceive the visual information in the a] video texts and b] still photo texts when processing and answering the comprehension questions to theses texts as indicated by the cued retrospective report?

2.1: Is there any correlation between the test takers' perceptions and their test scores?

Qualitative analysis of verbal data collected by the cued retrospective report revealed a matrix of perceptions related to the use of video and still photo texts in L2 listening test. Test takers provided a classification of some helpful as well as distracting aspects regarding both types of visual texts, supported by reasons explaining their perceptions.

Helpful aspects and reasons for video and still photo texts:

The magnitude of helpful aspects and reasons related to the video texts is larger than that with the still photo texts. A very important factor that distinguished video texts from still photo texts seems to be the positive effect of the presence of the speaker. Test takers perceived seeing the speaker's body language, face expressions, and kinesic behaviour to be helpful in linking information from the spoken input and the slides. This finding reflects the semantic congruity dimension of the visual taxonomy proposed by Suvorov (2013). Thus, implying that the video texts did achieve the goal set during designing these texts to be semantically congruent, in that visual and audio information complement and support each other. Semantic congruity was found in the results of other studies to be an effective factor in improving perceptions, though different terminologies were used, and this in fact adds to the validity of this dimension. Ginther (2002) found that when visuals contain information that complement the audio input, candidates' performance and positive perceptions towards that visual increase. The findings of the current study also revealed that seeing the speaker assist test takers' comprehension, help to stay focused, and to remember more information. These positive perceptions about the video texts indicate that the design and components of the video texts potentially elicited test takers' best performance, which is a goal of language testing stressed by Bachman and Palmer (1996). The fact that test takers reported benefitting from the speaker's movements in the video texts indicates that these movements of the speaker (e. g., kinesic behaviour and facial expressions, beside other factors like the content of the slides) probably facilitated employing certain metacognitive strategies, like assessment, by activating their topical and linguistic knowledge related to the video text, which can potentially help them to stay focused for longer time and comprehend larger proportion of the text. This effect is also reflected in the visual taxonomy (Suvorov, 2013) as the rhetorical effectiveness component, which means that test takers found the appearance of the speaker in the video texts to be persuasive and delivered information in an accessible way. This finding mirrors Wagners' (2006b) results, who also found that participants in his study reported the usefulness of seeing the speaker's face and gestures.

Another unique reason reported only with video texts is the perceived reality of the texts. Test takers perceived video texts as being more realistic than still photo texts and reflected real-life academic lectures. Hence, test takers could see the degree of correspondence between the components of the video text and the characteristics of the academic lectures, which is the specific TLU domain in this case. This indicates that the level of authenticity in video texts is largely higher than in the other texts- namely, still photos and audio texts. The different components that made the video texts, like the speaker, the slides, and the whole surrounding environment of a university classroom seems to have interacted together and created more representative output. Although it is not attainable to create the exact authentic, real-life situation in a test situation, nor it was the intend to do so, the employment of video texts had potentially reflected the expected level of difficulty that test takers will probably face in the real-life situations of the TLU domain (Field, 2013). In turn, this can potentially trigger cognitive processes that are similar to those normally used in the TLU (Field, 2012) and therefore achieve better level of cognitive validity of the test. If the right set of cognitive processes was employed, then test takers' performance in the test can be a good indicator of their abilities in that TLU, and that gives raise to the construct validity of the video texts as well, which reflects an important quality of language tests (Bachman and Palmer, 1996). All these outcomes of using video texts have the potential to minimize the use of test-wise strategies, and promote normal behaviour and strategies, which according to Field (2012) are signs of a cognitively valid test.

The use of slides was reported as a helpful aspect in both video and still photo texts, with a larger focus on them with the video texts. The beneficial role of the slides seems to be manifested in its connection with the speaker in the video texts. The speakers' occasional movements that refer to the slides seem to help link information from the spoken input and the visual cues in the slides like textual information or pictures and tables, and therefore complement each other (Ginther, 2002; Vandergrift and Tafaghodtari, 2010). This potentially can explain the raise in perceiving slides as helpful aspect with the video texts compared to still photo texts. Nevertheless, this cannot underestimate the usefulness of slides with the still photo texts as well. The content of the slides has potentially served to activate test takers' language knowledge and any related topical knowledge, which are vital processes in comprehending language input (Bachman and Palmer, 1996; Rost, 2013). The slides may also have helped test takers to associate its visual content with what they already have in their memory (Wagner, 2006). Also, the content of the slides may have assisted test takers to make better inferences about the text and test them against what is available in the slides by employ some metacognitive strategies like problem solving, which can result in better comprehension of the text (Vandergrift and Tafaghodtari, 2010), and better perceptions. Like the appearance of the speaker, using the slides can also be argued to increase the level of authenticity of the test as acknowledged by Field (2010), who explained that the use of Power Point slides in academic listening tests would reflect the real-life situation of university lectures. Lynch (2011) also called for incorporating multimodality in academic listening tests by making use of pictures and slides to reflect the type of listening normally encountered by learners in their academic life.

There were some useful aspects and reasons reported exclusively with still photo texts. In some instances, these perceptions completely contradict what has been stated as useful aspects and reasons with the video texts, which indicated that these references were made by different groups of test takers who perceived still photo texts in a different way from video texts. Test takers found still photo texts helpful because there is no distraction from the speaker, and therefore it is easy to ignore the whole screen, or to focus more on the slides zone only.

This result approves what Wagner (2006b) called for in his study, and that is the use of visual materials in listening tests can give test takers the option of where to devote their attentional resources. If they felt that they can perform better without looking to the visual materials, they can easily ignore them by looking away from the screen or even simply by closing their eyes. This might be related to test takers' different abilities in dealing with more than one modality at the same time. This result taps into a very important aspect discussed by Buck (2001) who argued that because people in general have different abilities of using visual information; it is important that L2 listening test designers focus only on assessing abilities related to process audio information only. His perspective is that utilizing video texts in listening tests would give a biased advantage to those test takers who can use the visual information in a skilful way over other test takers who are not as skilful. However, the results of the cued retrospective report had proved an alternative perception. Those test takers who potentially had some difficulties in processing the visual information were simply able to ignore the screen and focus only on the audio input. The use of visuals in listening test would possibly provide a wider base that suits the different abilities of people in processing different modalities than adhering to the use of audio texts only. It can be argued then using audio-only texts can be unfairly biased against listeners who are adept in using visual information, and it would underestimate their ability in processing the spoken input. As a result, their scores in such limited listening test would not be indicative to their real listening ability in the target language use domain. Logically, in today's lifestyle, the vast majority of people are living and interacting with various types of visual materials in almost all aspects of their lives. It seems to be that using audio-only listening tests would be a massive disadvantage for most test takers in our present time, and may result in a construct-underrepresentation case as explained by Messick (1996). That is, their performance on an audio only L2 listening test would not be indicative of their actual listening ability in a communicative language use domain. Instead of considering those listeners who are particularly adept at utilizing non-verbal information as being unfairly advantaged in a video listening test (as Buck seemed to imply), a more logical argument is that those listeners are unfairly disadvantaged in an audio-only listening test.

The design of the still photo texts that allowed for regular pausing time was perceived as helpful by some test takers. This in turn allowed those test takers to have time for writing their notes. Note taking is recognized as an important characteristic of the academic lectures (Bloomfield et al., 2010; Flowerdew, 1994). Therefore, the still photo texts can be said to facilitate this process during the test, even though it can be argued that the whole design of these texts does not reflect real-life situations, because of its stillness. However, considering that visual texts cannot be designed to exactly imitate real-life situations because that simply would not be possible in test situations, a reverse argument would be that these still photo texts provided test takers with the suitable environment to take notes, and therefore employ cognitive processes relative to those they would use in real-life academic listening.

All these components of the visual texts are related to the characteristics that are unique to listening and are in different ways representative to the academic listening target language use domain. Except for limited cases like listening to the radio or talking on the phone, in virtually all real-life listening situations the listener is able to see the speaker's kinesic behaviour and the surrounding physical setting. Elliot and Wilson (2013) argued that listening test design should always reflect TLU, therefore there is always strong ground for revising and expanding the listening construct to include visual materials that reflect the TLU domain.

In sum, the visual components in the video and still photo texts (with larger focus on video texts) proved to be to a large extent helpful to test takers, semantically congruent with the audio input, and rhetorically effective. According to the results, aspects of visual texts have the potential to rigger cognitive processes and consequently increase the level of authenticity of the L2 listening test. All these findings call for the need to expand the L2 academic construct to include visual materials.

Distracting aspects and reasons related to video and still photo texts:

Similar to the results of the perceived helpful aspects and reasons, some of the reported distracting aspect and reasons were shared in both the video and still photo texts. All the aspect and reasons stated to be distracting with the video texts were also reported with the still photo texts except the aspect of the speaker. It is found by some test takers that the speakers' movement is distracting because it drew their attention away from the audio input, which contradicts what was stated in the helpful aspects as the speaker found to be helpful by some other test takers. Likewise, test takers reported the concurrent display of visual with the audio input to be distracting in both the video and still photo texts because they found them to attract their attention away from the audio input. These results echo the findings in Suvorov's (2013) study, and supports Vandergrift's (2004) contention that processing visual materials might lead to limiting the capacity of working memory and result in distracting the listener from paying attention to the required source of information. However, this interpretation contradicts with most results of studies conducted on the effect of audio-visual condition on working memory processing. Many researchers found that the use of audio-visual presentations helped to enhance the encoding in the long-term memory and facilitated better processing in the shortterm memory (Mishra et al., 2013; Wiswede et al., 2007).

With both types of visual texts, test takers perceived the display of visuals in a listening test as a new practice which they are not used to in test situations. This came as surprising finding and it implied that test takers had a rooted presumption about the nature of a listening tests that employ audio materials only and came prepared to take a listening test in this traditional way. The presence of visuals confused the test takers because they could not apply the strategies they were taught to use in test preparation courses. This exact reason they stated is a clear prove of how test takers preferred to use test-wise strategies, which are believed to be a source of threat to the cognitive validity of listening tests (Field, 2012).

The distracting aspects and reasons reported with the still photo texts are revolving around the same centre, that is the lack of ongoing congruity between the visual and audio input. In other words, test takers found the still photo texts to be distracting because the speaker is not moving and the frequent pauses and changes of the photos at regular intervals, which consequently led to a kind of deficiency in the more normal flow of information, that is the continuous connection between the audio input and the picture. Because of the lack of synchronized communication between the speakers' on screen behaviour (e. g., pointing to a specific point on the slides) and what is said at a particular time, test takers did not find the visual materials in the still photo texts to be helpful nor convincing. The opposite condition was reported with video texts as finding them helpful and convincing because of the harmony between the speakers' behaviour and the slides and audio input. This reflects Ginther's (2002) results, as she found the aspect of semantic congruity to be significant to test takers and helps to improve their perceptions about the test as well as their performance. The results refer also to the importance of the dimensions of visual taxonomy (Suvorov, 2013). More specifically, according to this taxonomy, this type of visuals that is presented via the still photo condition characterizes a severe case in that taxonomy which is regarded as part of the visual language, and that is because it is sometimes found to be semantically incongruent, rhetorically ineffective, and liminal in its dynamism. In other words, it seems that the still photo condition in its current design, although perceived as more helpful than audio condition, did not reach to the validity level that allows it to be justifiably employed in L2 listening comprehension tests. This urges the need for further research with different designs of still photo texts in order to find out how these designs affect the performance of test takers. Overall, these results suggest that the decision to employ some type of visual by the test designers, either it is video or still photos, should be driven by the nature of the construct they wish to measure.

In sum, distracting aspects were found in both types of visual texts, with larger focus on still photo texts. The distraction as perceived by test takers can be rendered to the semantic incongruity, rhetorical ineffectiveness, and being liminal in its dynamism. There was also a preference for using test-wise strategies by test takers, as they perceived visual texts to be distracting because of not allowing them to apply these strategies.

Video texts proved to be superior to still photo texts in terms of helpfulness as perceived by test takers, and surpassed both still photo and audio texts in terms of test scores. General perceptions about the three text types by test takers are also in line with these results as revealed by the order of preference, with video texts in the first place, followed by still photo texts, and lastly comes the audio texts. The results of the present study support previous studies in the field of L2 listening tests with visual materials (Ockey, 2007; Suvorov, 2013; Wagner, 2006a, 2006b, 2010) and call for revising the construct of listening to include visual texts, especially video texts, in order to increase the level of authenticity of the test and represent the target language use domain.

5.3. Discussion of research question 3:

What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualized texts?

Results from the eye tracking revealed an interesting set of results. In general, test takers proved to fixate their eyes and spend total dwell time on video texts significantly more than still photo texts. No such difference was found with dwell rates, which implies that test takers revisited both types of visual texts to a similar extent. Moreover, fixation count according to AOI revealed that test takers focused more on the slides zone with the still photo texts, while similar amounts of fixations were recorded on both AOIs with the video texts. These results coincide with the results of the cued retrospective reports in this study, that is more attention was paid to the video texts compared to still photo texts as they fixated their eyes and spent longer watching time on them. This result also corresponds well with the way Holmiqvist and his colleagues (2011) defined fixation count as an indication of semantic informativeness and importance. It also matches previous results from usability research on fixations which indicates that viewers fixate their eyes more on objects and AOIs that they think are more significant than others (Poole et al., 2005). Hence, it can be argued that fixation count in the current study are more likely to be referring to the semantic importance of the video texts, rather than difficulties in processing the visual information in them.

In terms of the Total dwell time measure, test takers proved to spend longer time watching video texts than still photo texts as revealed by the eye tracker data. The results also positively correlate with test takers' perceptions about the video texts as helpful texts. Together, these results support the indication made by Holmiqvist et al., (2011) about interpreting longer measures of total dwell time as a reference to finding the object in question to be highly informative and interesting. This specific result also agrees with the eye-mind hypothesis by Just and Carpenter (1984), as test takers proved to have good correlation between their viewing patterns and their perceptions, which indicate that they were probably thinking about the object in focus during the time of watching it. The result also echo previous findings on the relationship between eye movements and positive perceptions (Suvorov, 2013).

Results of Dwell rate revealed no difference in the watching patterns between video and still photo texts. This result may indicate that test takers found both text types of relatively equal importance for the completion of the test task (Holmiqvist et al., 2011). The results of correlation between test takers' perceptions and their viewing patterns indicated a positive correlation between Dwell rates and perceiving still photo texts as helpful, which can be an indication that perceiving still photo texts as useful encouraged test takers to maintain a higher level of revisits to the screen. Also strong correlation was found between the total dwell time and perceiving video texts as helpful, which also may indicate that because test takers found video texts helpful, it encouraged them to revisit the screen more. However, no definite interpretation can be linked to this aspect, and that might be partially because of the novelty of establishing correlations between eye tracking measures and perceptions. This calls for further research with larger numbers of participants, maybe from different cultural and linguistic backgrounds, in order to establish more solid base for the relationship between perceptions and viewing patterns.In terms of test scores and the viewing patterns of test takers, no correlation was detected, suggesting that the nature and amount of test takers' interaction with the visual materials did not actually affect the test scores. It seems that there was a big variation in the amount of interacting with the visual materials by test takers who scored high as well as test takers who scored low. This result implies that test takers with different proficiency levels have benefitted from visual materials to different degrees. The fact that this was not reflected on their scores might be related to several factors. For instance, it might be the case that visual materials had helped some lower level test takers to understand the general idea of the text and that helped them to form positive perceptions about these materials; however, they had not provided the right answers to some questions that ask about specific details in the text.

Chapter Six Conclusion

L2 language testing has always been used as a tool that allows test developers, educators, and teachers to make inferences about the test takers' ability to use the language in situations other than the test which are called the target language use domains (Bachman and Palmer, 1996). These inferences are basically extracted from the candidates' performance (i. e., scores) in a specific L2 test. In theory, a language test should be representative to the situations of the specific target language use domain so the inferences based on that test can be valid. Looking at L2 listening for academic purposes, one can see that in almost all situations of this target language use domain the L2 listener has the opportunity to see the speaker and the surrounding physical situation where the speech takes place. Yet almost all L2 academic listening tests still use audio only texts to assess the test takers' ability to listen for academic purposes. Using the audio only texts in listening test was encouraged by some researchers like Buck (2002), although no empirical evidence was produced to support that claim. In essence, using audio only texts would prevent test takers to take advantages of different components available normally in real life academic listening situations like lectures or conference presentations, which would normally allow the listener to see the speaker's behaviour, gestures, and body language in addition to any supporting physical/visual materials they use like PPT slides or actual objects they might bring to the lecture for clarification purposes. Test takers would not normally be tested for their ability to listen to speech on the radio or over the phone, in which case including the visual aspect would be unsuitable. This would introduce what Messick (1996) explained as construct irrelevant variance, and it would threaten the validity of the test and the inferences made on its bases. However, because the target language use domain, in this case academic listening, normally involves different types of visuals, like the lecturers, the physical arrangement of the class rooms, the supporting visual materials of the lectures, and many more, it would be threatening to the validity of the test if an attempt was made to exclude all these variable from the test, and that would be a case of construct underrepresentation as described by Messick (1996).

Consequently, the main purpose of this study was to explore the effects of implementing two types of visual materials: video and still photo texts, on the performance of test takers in L2 academic listening comprehension tests. The motivation behind this study was multi-fold. Theoretically, there is a tendency among some researchers to advocate the idea of including visuals in L2 listening comprehension tests, suggesting that it is theoretically sound (Baltova, 1994; Dunkel, 1991; Progosh, 1996; Rost, 2006). However, the empirical research did not present consistent solid results on which a radical change to the current design of listening tests can be based (Buck, 2001; Gruba, 1993; Wagner, 2006b, 2010). The effect of visual materials on the test takers' performance in many studies was not clear and in some cases rather confusing. In addition, given that most of the research on visuals investigated the use of video texts only, it seemed imperative to include the use of still photo texts as well since it is the only form of visuals currently in use in some TOEFL CBT, and that is only for saving the face validity of the test, with no empirical evidence provided to support it.

In order to achieve the purpose of this study, a mixed method was applied, based on the triangulation design of Creswell and Plano Clark's (2011). This triangulation design required the collection of three sets of data: a) the quantitative test scores data, b) the qualitative cued retrospective report data, and c) the quantitative eye tracking data.
Accordingly, these sets of data produced three main types of results: first, test takers' performance in the three text types of the test (i. e., video, still photos, and audio), which proved by the results to be significantly higher with video texts compared to audio texts. Nevertheless, no other significant differences in their performance was detected (i. e., not between the video and still photo texts nor between the still photo texts and the audio texts).
Second, test takers' perceptions about the two visual texts, where they perceived the video texts to be more helpful than still photo texts in both processing the texts and answering the test questions. Lastly, test takers' viewing behaviour which revealed that test takers did spent longer time watching video texts, and also had more fixations on video texts. Hence, these sets of data highlighted the importance of using the eye tracking tool and the cued retrospective report jointly as supporting materials, which both provided the researcher with the opportunity to explore the deeper effects of including visual materials in L2 listening comprehension tests. Relying on the test takers' scores only would have misled

the inferences made about the effects of visuals, because the scores would normally give the final outcome of the test without exploring the process that the test takers went through during processing the test materials and answering the questions. These results have generated some suggestions and implications on the design of L2 academic listening tests as will be discussed in the following sections of this chapter. Most importantly, the inclusion of visual materials- particularly video texts- should be considered in order to reflect the target language used domain. This calls for the revision and expansion of the current L2 listening construct that employs audio materials only.

In this concluding chapter, a summary of the results of the three research questions is provided. Then an outline of the theoretical, methodological, and testing implications that resulted from the study is presented. Lastly, a discussion of suggested guidelines for future research is given.

6.1. Summary of the main findings

The study was built on three main research questions. To address the first question "To what extent does the performance of L2 test takers differ on a listening test with a] video texts, b] still photo texts, and c] audio-only texts?", a within-subjects design was used to examine the difference in test takers' performance (*n* = 30). Each test-taker was examined with a total of six listening texts (i. e., 2 video texts, 2 still photo texts, and 2 audio texts). To determine if there was a difference in the performance in the three different text types; one way within-subjects -repeated measures- ANOVA was conducted. The results of this

analysis indicated that the test takers' performance in the video texts was significantly higher than their performance in the audio texts at *p* = .03. Meanwhile, no significant difference was detected between their performance in the still photo texts and audio texts, nor between the video texts and the still photo texts. This result implies that the use of video texts might have positively affected the performance of the test takers compared to the traditional way of using audio-only texts in the L2 listening comprehension tests. These results in fact go in line with the results of previous studies of Baltova (1994), Suvorov (2013) and Wagner (2006b), who similarly found that higher scores were gained by test takers in video supported listening test compared to test takers in an audio-only listening test.

The second research question consisted of two parts, the first part asked, "How do L2 test takers perceive the visual information in the a] video texts and b] still photo texts when processing and answering the comprehension questions to theses texts as indicated by the cued retrospective report?". Followed by the second part "Is there any correlation between the test takers' perceptions and their scores?". In the first part of the question, test takers' perceptions about the video and still photo texts while processing them were extracted using heat maps as cues during the retrospective verbal report. The qualitative analysis of the data revealed that helpful and distracting aspects were found by the test takers in both video and still photo texts but at different degrees. Regarding the helpful aspects with the video texts, the aspect of the speaker was perceived as highly helpful by test takers, and that led to providing more details on the helpful sub-aspects related to that main one, like those related to the speakers' facial expressions (e. g., lip movement, eye contact with the camera, etc.) which were focused on by 50% of the total number of test takers. Also the speakers' body language was mentioned by 20% of the test takers as helpful feature, in addition to the speakers' general character like their appearance and engagement with the text which was stated by 37%. There was also another helpful aspect related to the video texts, which is the slides' aspect. Although test takers had focused to a greater degree on the speaker related aspect regarding the video texts; the lecture related aspect (slides) had also gained some attention, with 33% of the participants referred to helpfulness of the textual information provided in the slides, and 23% of them mentioned the usefulness of the pictures and diagrams that were available in the slides in boosting their comprehension of the text. Comparing this to the helpful aspects in the still photo texts, it is found that there was a complete absence of the role of the speaker as helpful aspect in these texts. The only mentioned helpful aspects were related to the design of the lecture, like the Power Point slides (43%) and the nature of displaying the still photo texts that allowed for pausing times (17%) which mostly helped the test takers in taking note.

In terms of the reasons for perceiving the aspects in the video texts as helpful, it was found that test takers focused on these texts due to three main reasons, which are speaker related reasons, slides related reasons, and the last reason which is the general need for visuals. Reasons related to the speaker included: general face expressions that helped comprehension (30%), gestures and body language that helped test takers to connect audio to textual information in the slides (27%), and lastly and most importantly as it seems that the presence of the speaker helped them to focus more on the lecture (57%). Regarding the reasons related to the slides, test takers found that the textual information on the slides helped to confirm what the speaker said (40%), and the diagrams and pictures helped to summarize the main idea of the text (37%). Finally, test takers revealed that they found video texts helpful because they needed to see visuals while the audio text was playing, and they found that to be boosting their comprehension and also helping to remember information in a better way (53%), and the video also reflected real situation which was like a real lecture (50%). With respect to the still photo texts, the test takers explained four main reasons, with the first was that the still photo texts provided less distraction in the part of the speakers because they were not moving which helped them to put more focus on the slides and the audio texts (17%), also these texts helped to summarize the situation of the lecture (27%), gave them more time to write notes (23%), and because this type of visual is still, it was easy to ignore if they felt the need to do that (13%).

In addition to the mentioned helpful aspects and reasons that are related to both video and still photo texts, test takers also found some distracting aspects, and reasons in these visual texts. In relation to the aspects related to the video texts, 17% of the test takers found that the movement of the speaker is distracting, and 13% stated that the concurrent presentation of the audio and video stimuli was overwhelming to them. With regard to the still photo texts, test takers found that the visual design of the texts where the pictures change at regular intervals was very distracting (33%), and like with the video texts, 27% found the concurrent audio/visual stimuli in the still photo texts to be distracting as well.

The reasons for finding video and still photo texts distracting as stated by the test takers were rather different in their magnitude, with only two reasons related to the video texts compared to five with the still photo texts. 17% of test takers found that video texts were distracting because they attracted their attention to the extent that led them to miss parts from the spoken input and interfered with their note taking process. Also, 30% stated that taking a listening test with video is a new technique for them which they were not trained for during their language test preparation courses. In terms of the still photo texts, the test takers claimed that these texts were distracting because its design of frequent pauses was strange and distracting to them (37%), no clue can be extracted from the speaker (33%), they couldn’t build a link between the audio input and the photo on screen

(17%), and similar to the video texts, still photo texts led some test takers to pay excessive attention to the screen and miss parts from the audio (23%), and once more, 33% viewed the still photo texts as a new and strange type of listening that they were not prepared to have in a listening test.

In addition to the test takers' perceptions on how these visuals affected the way they processed the texts, they were also asked to indicate whether the visuals helped them to answer the test questions. The results revealed that the test takers perceived video texts to be more helpful in answering the test questions than still photo texts, with marginally significant difference (*p* = .06). Test takers were also asked to comment on their general perceptions about the three types of texts and to rank them according to their preference. The results showed that video texts occupied the first place as 53% of the total number of test takers perceived it to be the most helpful, secondly still photo texts at 27%, and lastly audio texts at 20%.

The second part of research question two investigated whether there was a connection between test takers scores in the L2 listening test and their perceptions on the two visual texts. Spearman rank-order correlation test was conducted, and the results revealed a significant positive relationship between video texts scores and the test takers' perceptions of these texts as being helpful, at *p* = .01. However, when performed with the still-photo texts, the correlation results revealed positive but insignificant relationship between the test-taker' scores and their perceptions of these texts as helpful at *p* = .24.

In response to the third research question "What kinds of viewing patterns can be observed across the video and still photo texts as indicated by the eye tracker? Is there a connection between these patterns and a] test takers' performance in the test, and b] their perception of the visualized texts?", paired samples *t*-tests were performed with three eye tracker measures (i. e., fixation count, dwell rate, and total dwell time) for the video and still photo texts. The results revealed that test takers fixated their eyes on the video texts significantly more than the still photo texts at *p* = .01, and spent longer total time watching the video texts than still photo texts at *p* = .00. Meanwhile, the difference in the revisits rate (dwell rate) to the two visual texts was not statistically significant. In addition, the analysis of fixation counts per AOI in the two visual texts revealed that test-taker treated the speaker's AOI and the slides' AOI similarly with the video texts, however, with still photo texts they fixated their eyes more on the slides' AOI than on the speaker's AOI.

To answer the second part of this research question, Pearson correlation coefficient was used and it showed no statistically significant relationship between the test takers' viewing patterns and their test scores. However, when the correlation test was performed with the test takers' perceptions on the two visual text types, the results showed a marginally significant relationship between the number of revisits to the screen (dwell rate) and the test takers' perception of the still phot texts as useful at p = .04. Moreover, a statistically significant relationship was found between the total amount of time the test takers spent watching the video texts and their perceptions of these texts as useful at p = .04.

6.2. Implications of the study

The findings of the study yielded some theoretical, methodological and testing implications concerning L2 listening assessments and ability. Theoretical implications relate to the production of a modified construct for L2 listening comprehension that suits the future generations of L2 test takers. Methodological implications discuss the practicality of using eye tracking technology in L2 listening research. In addition, testing implications discuss the usability of video and still photo texts in L2 listening comprehension tests.

Theoretical implications

The findings of this study present evidence in support for the use of visual materials, specifically video texts. Test takers have scored higher in video texts compared to the still photo texts and audio texts. They also perceived video texts to be the most helpful and most natural type of input, and spent longer time watching them in comparison with still photo texts. This implies that delivering the test content using both the visual and audio channels can contribute to the construct validity of the L2 academic listening test. It has been maintained that the essence of testing the listening ability is to test the characteristics that are exclusive to listening and not to include any other outer characteristics (Buck, 2001; Rost, 2006). In the case of academic listening, and in fact most real-life listening situations, the target language use domain involves more than the audio channel. Listeners normally have the opportunity to see the speaker in almost all cases of listening, and that include seeing the face expressions, the body language, the surrounding context, etc. Hence, the visual side is an integral part of the listening experience, and if test designers seek to design a test that includes the characteristics which are unique to listening, then they should include the visual aspect in the listening construct and integrate visuals into the design of their tests since the target language use domain involves the ability to use both the audio and the visual information.

Methodological implications

The use of cued retrospective reports in this study provided the chance to systematically investigate test takers' watching behavior without interrupting them during the test. The heat maps gave an instant clue to the test takers to remember the way they viewed the screen with each visual text. Cued retrospective reports had also prevented the possibility of losing parts of information due to forgetting how exactly they behaved when looking at the visuals in case they had not seen the clues with the heat maps.

Another methodological implication which is considered vital is the complementary nature of this study. The three major parts that constituted the study, namely- test scores, verbal report, and the eye tracker data, served to provide larger and more comprehensive picture of the process of test taking by covering both the product of the listening test and the process that led to that product.

Testing implications

The results of this study have also some implication for the assessment of L2 academic listening comprehension. As seen from the results of repeated measures ANOVA

in research question one when comparing the scores of the three text types, test takers' performance with the video texts was significantly higher than their performance with the audio texts, which implies that the video texts might have a facilitative effect on test takers' performance. In addition, although it was not proved that the still photo texts had a facilitative impact on test takers' performance, it was not proved that they have a harmful effect either. Following Ginther's (2002) call that if visuals had no harmful effect on test takers performance then they should be included in the listening tests, this study therefore suggests that both video and still photo text should be taken into consideration by test designers when they attempt to develop L2 academic listening comprehension tests.

In addition, evidence from the eye tracking data revealed that the test takers fixated their eyes more on the video texts and spent longer time watching them too. Linking that to the test takers' overall perceptions about the three types of texts used in the study, the verbal report data showed that more than half of them (53%) perceived video texts to be the most helpful compared to the other two types of texts. Although this study is devoted to the assessment of listening comprehension, hence it has more implications to the development of listening tests, these results can also serve a pedagogical purpose. The results show that the visual information is very important for processing spoken texts and utilizing this type of information is vital to the skill of L2 listening, therefore they should be used in the listening courses by language teachers at a wider scale to provide more support to L2 language learners. Lastly, the simultaneous use of acoustic and visual input in this study implies that the term Listening Test might not be adequately suitable for this type of tests, and new terms should be considered in order to reflect the nature of the test.

6.3. Limitations of the study

This study has a number of limitations that need to be acknowledged. Basically, the limitations are linked to some issues with the use of eye the tracking device, and also to the implementation of verbal reports, and finally discussing the generalizability issue.

Eye tracking:

Although the use of eye tracking added a very important dimension to the study, the type of equipment used in this study has a limitation that might have affected test takers performance. In particular, the eye tracking data were collected via EyeLink 1000, desktop mount device (as shown in figure 3.4), which required placing the participant's head on the chin and forehead rest for the whole test period which lasted around 47 minutes. In some cases, this had resulted in the test-taker being disturbed and not feeling comfortable at the end of test. The device might have also affected the quality of calibration in the case of this test, although the calibration sample was high (1000 Hz), which largely surpasses what is recommended by Jacob and Karn (2003) that the sampling rate should be at least 250 Hz in order to collect accurate data. The quality of calibration might be affected when test takers move their heads away from the desktop mount in order to take notes and then look again at the screen. The eye tracker at this point loses the participants' eye and then it has to search again for their fovea in order to resume recording the eye position. During this stage the eye might not be detected easily and it takes the eye tracker some time to recognise the test takers' fovea again and start recording the viewing patterns at 1000 HZ again. This in fact had also affected some of the test takers watching records by missing few seconds of actual watching that were not recorded. Therefore, it can be assumed that the act of taking notes might had affected the eye tracking results, and that was not considered during the analysis of data.

Another limitation related to the eye tracker is the process of defining static AOIs in the video texts. Static AOIs were applied to both video and still photo texts in order to be systematic and to collect data that can be compared consistently. However, dynamic AOIs with video texts can give more accurate information about the fixation count in the Speakers' AOI as they move their hands toward the slides from time to time. This fact may have resulted in the case of static AOI to collecting fixation points on the slides' zone, where they were actually related to the speakers' hands. Although caution had been taken while recording the video texts to manage the position of the speaker in one side, and the AOI was defined very carefully to include the speakers' zone, it was not possible to control the overlaps between the speaker and the slides in the video texts with a static AOI.

Verbal report in English:

The nature of this type of qualitative data- the cued retrospective report- required extracting verbal information from the test takers. Because in this study the participants were from two different proficiency levels, namely advanced and intermediate, the quality of the verbal data was affected. In the case of participants with intermediate level of English proficiency, the verbal interaction in a number of instances was limited because some of them could not express their thoughts clearly in English and kept their answers to the minimum. As a result, the nature of information collected from those cases was rather shallow. It would be suggested that researchers would benefit a lot more if verbal data was collected in the test takers' native language if possible. Only the few cases with intermediate level Arabic participants were done in Arabic and that resulted in much longer and more detailed reports, while that was not possible with the rest of intermediate level test takers whose native tongues are different. That had possibly affected the quality of the verbal information as it was noticed by the researcher that in few instances those lower level test takers were providing very short answers (e. g., yes, no) just to finish the interview quickly and not be forced to speak more.

Generalisability:

The total number of participants in this study is 30, with intermediate and advanced learners of English. This number of participants is relatively small for the results to be generalizable to the wider population of L2 test takers. In addition to the limited sample size, the number of visual texts watched by each test-taker were also limited (2 texts per visual type). Although this was decided for practicality reasons, as the test takers in the pilot study were very tired and bored at the end of the test that consisted of 9 texts instead of 6, but it resulted in less total amount of watching for the video and still photo texts, which might have affected the reasonability of the results. It would be advised that future research might consider either including larger numbers of participants to increase the validity of scores, perceptions, and watching patterns' comparison, or increasing the number of texts that can be presented along two or three test sessions instead of one long test to minimise the pressure placed on test takers.

More importantly, in controlled laboratory experiments, there is always some degree of uncertainty to whether the participants' behaviour and performance in that controlled environment really reflects their performance in the real-life situations and domains in which they are going to communicate. Therefore, even with larger numbers of participants, laboratory experiments always remain partially unclear in terms of the extent of their generalisability.

6.4. Suggestions for future research

This study, with its results, implications and limitations has brought up some directions and suggestions for future research. These suggestions can be organized in three main groups that comply with the three major areas in this research- namely, suggestions on the use of visuals in listening tests and how it can affect test performance, suggestions on improving the way cued retrospective reports can be used in the future, and lastly suggestion on the use of eye tracking device in listening research.

Comparative research on visual listening comprehension tests:

Research on the use of visual texts in listening comprehension tests is very limited in general. Most of this research was devoted to the impact of using video texts on the performance of test takers (Baltova, 1994; Başal, Gülözer, and Demir, 2015; Batty, 2015; Coniam, 2001; Gruba, 1993; Suvorov, 2013; Wagner, 2006b, 2010). However, comparative studies of including both video and still photo texts are very scarce (Ockey, 2007; Suvorov, 2009) and the evidence brought by these few studies is very vague in terms of how the videos and still photos affected the test takers' performance. On this base, there seems to be a great need for more in depth research that compares the impact of using these two types of visuals on the listener's performance. More precisely, considering the fact that the current research was conducted with a limited number of participants, there appears to be a need for more comparative studies in order to see whether test takers' performance really differ with different modalities (i. e., video, still photos, and audio-only), taking into account other variables such as cultural background of participants and types of questions.

In this study the participants were from diverse cultural and linguistic backgrounds, with a dominating number of Chinese participants who, mostly, received IELTS preparation courses and were in many cases as proven by the results of the retrospective report employing test wise strategies which they will properly not be using in the real-life situations of L2 listening. A useful future study might be one which compares the performance of test takers from two specific backgrounds, with a detailed investigation of their educational backgrounds in order to address whether this variable can affect their performance in the different modalities of the test.

It would also be useful to employ more types of questions in the L2 academic listening test other than multiple-choice questions and gap filling, like for instance writing a summary of the lecture or providing an extended written response. The choice of the type of questions depends largely on the specific target language use domain and should reflect the characteristics of that domain.

In addition, investigating the effects of other visual texts with different types of stimuli, for instances, including communicative situations like short conversations or group discussions would be useful. This type of research could potentially help to explore the functions that different visuals may play in L2 listening comprehension tests. In 2002, Ginther conducted a study to examine the effects of the presence or absence of content and context still photo(s) with three different stimuli types: dialogues, academic discussions and mini talks. The results showed that the still photos only facilitated test takers' performance when they conveyed information that complement the audio input. It would be convenient that future studies replicate this type of research with video texts, to inspect how it would affect test takers' performance compared with their performance on the still-photo texts.

The use of cued retrospective reports in visual listening tests

The use of this type of verbal reports has resulted in some suggestions for future research to add to the quality and benefits of using cued retrospective reports. First, future studies might consider conducting the cued retrospective reports immediately after each visual text. In other words, it would be useful to show the participants their viewing patterns via the heat maps after each text, and not waiting until they finish the whole test. However, researchers should realise that executing the heat maps in this way may require some technical adjustments to allow generating a separate session of data viewing for each individual trial. That might be time consuming, but it is believed that its advantages can surpass the disadvantages in this respect. That would possibly increase the benefits of the heat maps by allowing the test takers to recall even more information about the way they processed the text and how they answered the test questions with relation to the visual, and this would result in highly reliable verbalisation of data that can make it possible to more accurately mirror the cognitive processes employed by the test takers.

Another suggestion for future studies is to conduct the verbal reports in the test takers' native language if possible. As seen from the results of this study that in some instances, the test takers provided monosyllabic responses (yes/no) and couldn’t give more details about certain aspects because of their relatively low speaking proficiency.

Conducting the verbal reports in participant’s native languages, although it can be challenging to find interpreters for specific languages, can potentially encourage the participants to express their ideas and perceptions in more depth.

The use of eye tracking technology

The eye tracker is a robust device that can open the door to the investigation of variant methods of L2 language tests that employ visuals. For future research, it would be useful to investigate the viewing patterns of test takers while answering the test questions, and not only while watching the visual texts. This can be achieved by providing the test items simultaneously with the visual text in one split-screen. Examining how the test takers interact between the visual and the test items might be informative and could reveal more details about the extent to which different visuals can affect performance on different types of questions. For instance, multiple-choice questions might prompt a viewing pattern that is different from the patterns that could be associated with gap filling questions.

Eye tracker can be also used to inspect the effects of listening with other types of support. For instance, it can be used to find out more about the impact of listening to a text twice and how that can affect the test takers performance and viewing behaviour. Researchers can compare how test takers approach the test items when they listen once to a text and whether they use different viewing behaviours when they listen twice, which can reveal a lot about the cognitive processes employed with each technique.

Another potential use of eye tracker is that; it can be used to find out how listeners who are allowed to preview test items before listening are attending to the test items and compare that to listeners who are not allowed to preview the items prior listening to the text. Again, that can assist researcher to look into the cognitive processes in the two cases and inspect any potential difference.

6.5. Conclusion

The present thesis introduced a modern technique in order to shed more light on the process of interacting with visuals during L2 listening comprehension tests. The mixed method approach applied in this study with the triangulation of quantitative and qualitative data allowed for better and deeper understanding of the way test takers tackled the listening texts and test questions, the reasons of doing so and their perceptions, and the extent to which visual information helped them to process the texts and answer the test items. This approach resulted in minimising the reliance on inferences about the test takers' listening abilities based only on test scores as the way traditional research on listening tests was conducted. It is believed to be the first research that employs eye tracking technology to compare the impact of including two important types of visuals, namely video and still photos, on test takers' performance in L2 academic listening comprehension tests.

The results demonstrated how the two types of visuals were perceived and viewed differently by test takers, although this difference was not reflected in their scores in the video and still photo texts. Hence, the use of cued retrospective reports and the eye tracker managed to reveal a deeper level of variation among test takers' performance. It is believed that the results of this research, notwithstanding its limitations, presents significant theoretical and practical contributions to the use of visuals on L2 listening comprehension tests.

Appendices

Appendix 1

Transcripts of lectures and their test questions used in the study.

1- Lecture: Architecture

Good morning. In the last few lectures I've been talking about the history of domestic building construction. But today I want to begin looking at some contemporary, experimental designs for housing. So, I'm going to start with a house which is constructed more or less under the ground. And one of the interesting things about this project is project is that the owners- both professionals but not architects- wanted to be closely involved, so they decided to manage the project themselves. Their chief aim was to create somewhere that was as environmentally-friendly as possible. But at the same time they wanted to live somewhere peaceful – they both grown up in a rural area and disliked urban life.

So the first thing they did was to look for a site. And they found a disused stone quarry in a beautiful area. The price was relatively low, and they liked the idea of recycling land, as it were. As it was, the quarry was an ugly blot on the landscape, and it wasn't productive any longer, either.

They consulted various architects and looked at a number of designs before finally deciding on one. As I've said, it was a design for a sort of underground house, and it was built into the earth itself, with two storey. The north, east and west sides were set in the earth, and only the sloping, south-facing side was exposed to light. That was made of a double layer of very strong glass. There were also photovoltaic tiles fixed on the top and bottom of this sloping wall. These are tiles that are designed to store energy from the sun. And the walls had a layer of foam around them too, to increase the insulation.

Now, what is of interest to us about this project is the features which make the building energy-efficient. Sunlight floods in through the glass wall, and to maximize it there are lots of mirrors and windows inside the house. That helps to spread the light around. So that's the first thing – light is utilized as fully as possible.

In addition, the special tiles on the outside convert energy from the sun and generate some of the house's electricity. In fact, and it's possible that in future the house may even generate an electricity surplus, and that the owners will be able to sell some to the national grid.

As well as that, wherever possible, recycled materials have been used. For example, the floors are made of reclaimed wood. And the owners haven't bought a single item of new furniture – they just kept what they already had. And then there is the system for dealing with the waste produced in the house. This is dealt with organically – it's purified by being filtered through reed beds which have been planted for that purpose in the garden. So the occupants of the house won't pollute the land or use any damaging chemicals.

It's true that the actual construction of the house was harmful to the environment, mainly because they had to use massive amounts of concrete – one of the biggest sources of carbon dioxide in manufacturing. And, as you know, this is very damaging to the environment. In total, the house construction has released 70 tons of carbon dioxide into the air. Now that's frightening thought. However, once the initial 'debt' has been cleared – and it's been calculated the this will only take fifteen years- this underground house won't cost anything – environmentally I mean – because unlike ordinary houses, it is run in a way that is completely environmentally friendly.

So, eco-housing like this is likely to become much more…

Architecture questions

Chose the correct letter, **A**, **B** or **C**.

Q1] - The owners of the underground house

A- had no experience in living in a rural area.
B- were interested in environmental issues.
C- wanted a professional project manager.

Q2]- What does the speaker say about the site of the house?

A- The land was quite cheap.
B- Stone was being extracted nearby.
C- It was in completely unspoilt area.

Complete the notice below. Write **ONE WORD ONLY** for each answer.

Design:

- The south facing side was constructed of two layers of Q3] ---------------------.

Special features:

- The system for processing domestic Q4] -------------------- is organic.

Environmental issues:

- The use of large quantities of Q5] ------------------- in construction was environmentally harmful.

2- Lecture: Health

Good morning, good morning, everyone, and welcome to our regular lecture on health issues. This series of lectures is organised by the Students' Union and is part of the union's attempt to help you, the students of this university, to stay healthy while coping with study and social life at the same time.

Now, stresses at university, being away from home and having to look after yourselves, learning your way around the campus all contribute to making it quite hard sometimes to ensure that your diet is adequate. So today I'm going to talk about ways of making sure that you eat well while at the same time staying within your budget.

If you have a well-balanced diet, then you should be getting all the vitamins that you need for normal daily living. However sometimes we think we're eating the right foods but the vitamins are escaping, perhaps as a result of cooking and anyway we're not getting the full benefit of them. Now, if you lack vitamins in any way the solution isn't to rush off and take vitamin pills, though they can sometimes help. No it's far better to look at your diet and how you prepare your food.

So what are vitamins? Well, the dictionary tells us they are food factors essential in small quantities to maintain life. Now, there are fat soluble vitamins which can be stored for quite some time by the body and there are water soluble vitamins which are removed more rapidly from the body and so a regular daily intake of these ones is needed.

OK, so how can you ensure that your diet contains enough of the vitamins you need? Well, first of all, you may have to establish some new eating habits! No more chips at the uni canteen, I'm afraid! Now firstly, you must eat a variety of foods. Then you need to ensure that you eat at least four servings of fruit and vegetables daily. Now you'll need to shop two or three times a week to make sure that they're fresh, and store your vegetables in the fridge or in a cool dark place. Now let's just refresh our memories by looking at the Healthy Diet Pyramid. OK, can you all see that? Good.

Well, now, as you see we've got three levels to our pyramid. At the top in the smallest area are the things which we should really be trying to avoid as much as possible. Things like Example yes, sugar, salt, butter all that sort of thing.

Next, on the middle of our pyramid we find the things that we can eat in moderation. Not too much though! And that is where we find milk, lean meat, fish, nuts, eggs. And then at the bottom of the pyramid are the things that you can eat lots of! Because they're the things that are really good for you and here we have bread, vegetables and fruit. So don't lose sight of your healthy diet pyramid when you do your shopping.

Health questions

(1) Choose the correct answer.

According to the speaker:

(Q1) this lecture is about

A- Campus food. C- Sensible eating.

B- Dieting. D- Saving money.

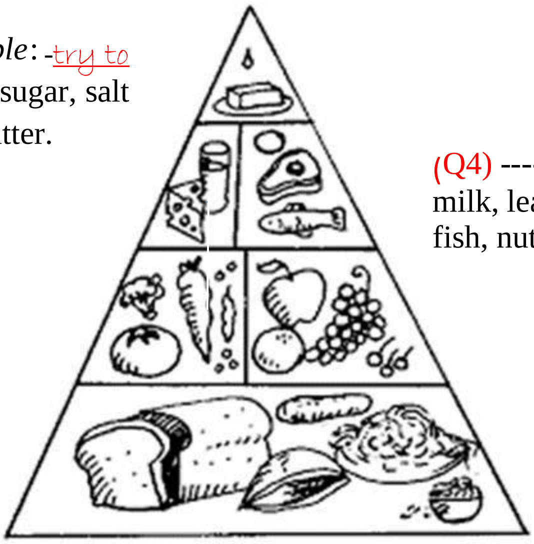
(2) Complete the notes. Write **NO MORE THAN THREE WORDS** for each answer.

A balanced diet.

(Q2) A balanced diet will give you enough vitamins for normal daily living. Vitamins in food can be lost through --------------------.

(Q3) Buy plenty of vegetables and store then in ----------------------------------.

(3) Complete the diagram by writing No MORE THAN THREE WORDS in the boxes provided.



Example: -try to avoid sugar, salt and butter.

(Q4) ------------ milk, lean meat, fish, nuts, eggs.

(Q5) ------------------ Bread, vegetables and fruits.

3- Lecturer: Sports

Good morning and welcome to the University's Open Day and to our mini-lecture from the Sports Studies department. Now the purpose of this lecture is twofold: one — we want you to experience a university lecture, to give you a taste of what listening to a university lecture is like, and two — we want you to find out something about the Sports Studies program at this university. So feel free to ask any questions during the talk and I'll do my best to answer them. Right — so what does a course in Sports Studies involve? Well, you wouldn't be blamed for not knowing the answer to this question because Sports Studies as a discipline is still comparatively new. But it's a growing area and one which is now firmly established at our university.

Now there are three distinct strands to Sports Studies and you would need to choose fairly early on just which direction you wanted to follow. And I'll just run over these now. Firstly, we've got the Sports Psychology strand, secondly, we've got the Sports Management strand, and last, but not least, there's the Sports Physiology strand. So Just to recap there's Sports Psychology, Sports Management, and Sports Physiology.

Let's look first at Psychology. Now the people who study Sports Psych want to work with top athletes, and they're looking at what will take those athletes that one percent extra. What makes them win? When all other things are equal, physically all other things are equal, they want to know … what are the mental factors involved? The Sports Psychologist works closely with the athlete through his or her training program and becomes an integral part of the team. In fact you could say that they play just as important a role as the coach. So if you're interested in what makes people win this could be the area for you.

Now secondly, we've got the strand which I referred to as Sports Management and this goes hand in hand with the area of Sports Marketing. So you might like to think of this area as having two branches: Management and Marketing. On the Management side we look at issues relating to the running of sports clubs, management of athletes that sort of thing. But then on the other side, we've got Sports Marketing. And this is the side that interests me more because here we will look at the market forces behind sport. Questions like: why do people spend their money on a football match, or a tennis game rather, than say on buying a CD or going to the cinema? What are those market forces?

Sport used to just compete with sport. Nowadays it competes with other leisure activities. The spectators go to sport to be entertained rather than out of loyalty to a team. They want to have an evening out and they don't want the cheap seats any more they want good seats they want entertainment. And the professional sportsmen and women respond to this without question. They're there to give a performance. They provide the entertainment. So in the marketing course we address all these commercial issues and we look at how this hooks back into the Management of sport.

Now the third branch of Sports Studies sometimes comes under another name and is also known as Exercise Science. And again here we find that there are two distinct types of exercise science. The first is working very much at the macro level. What I call the huffing and puffing people. So this looks at fitness testing, body measurements, all that sort of thing. But the more interesting side of sports Q41 physiology, at least in my view, is the side that looks at the micro level, looking at cellular change. They're doing cellular research, looking at changes in body cells when the body is under stress.

So that just about brings us to the end of our mini-lecture for today. I hope you've found it interesting and I look forward to seeing you all on our course next year. Feel free to come and talk to me if you want any more information. I'll be over at that notice board near the main entrance.

Sports questions

Complete the lecture notes. **Use NO MORE THAN THREE WORDS** for each answer.

- The purpose of the mini lecture is to experience Q1] ------------------------, and to find out about Q2] ---------------------.
- The three strands of Sports Studies are: a- Sports psychology b- Sports Q3] ------------------.

c- Sports physiology.

- The psychologists work with Q4] ------------------.

- Sports now competing with Q5] --------------------.

4- Lecture: social science

Today, I'm going to be talking about time. Specifically I'll be looking at how people think about time, and how these time perspectives structure our lives. According to social psychologists, there are six ways of thinking about time, which are called personal time zones.

The first two are based in the past. Past positive thinkers spend most of their time in a state of nostalgia, fondly remembering moments such as birthdays, marriages and important achievements in their life. These are the kinds of people who keep family records, books and photo albums. People living in the past negative time zone are also absorbed by earlier times, but they focus on all the bad things – regrets, failures, poor decisions. They spend a lot of time thinking about how life could have been.

Then, we have people who live in the present. Present hedonists are driven by pleasure and immediate sensation. Their life motto is to have a good time and avoid pain. Present fatalists live in the moment too, but they believe this moment is the product of circumstances entirely beyond their control; it's their fate. Whether it's poverty, religion or society itself, something stops these people from believing they can play a role in changing their outcomes in life. Life simply "is" and that's that.

Looking at the future time zone, we can see that people classified as future active are the planners and go-getters. They work rather than play and resist temptation. Decisions are made based on potential consequences, not on the experience itself. A second future-orientated perspective, future fatalistic, is driven by the certainty of life after death and some kind of a judgement day when they will be assessed on how virtuously they have lived and what success they have had in their lives.

Okay, let's move on. You might ask "how do these time zones affect our lives?" Well, let's start at the beginning. Everyone is brought into this world as a present hedonist. No exceptions. Our initial needs and demands – to be warm, secure, fed and watered – all stem from the present moment. But things change when we enter formal education – we're taught to stop existing in the moment and to begin thinking about future outcomes.

But, did you know that every nine seconds a child in the USA drops out of school? For boys, the rate is much higher than for girls. We could easily say "Ah, well, boys just aren't as bright as girls" but the evidence doesn't support this. A recent study states that boys in America, by the age of twenty one, have spent 10,000 hours playing video games. The research suggests that they'll never fit in the traditional classroom because these boys require a situation where they have the ability to manage their own learning environment.

Now, let's look at the way we do prevention education. All prevention education is aimed at a future time zone. We say "don't smoke or you'll get cancer", "get good grades or you won't get a good job". But with present-orientated kids that just doesn't work. Although they understand the potentially negative consequences of their actions, they persist with the behaviour because they're not living for the future; they're in the moment right now. We can't use logic and it's no use reminding them of potential fall-out from their decisions or previous errors of judgment – we've got to get in their minds just as they're about to make a choice.

Time perspectives make a big difference in how we value and use our time. When Americans are asked how busy they are, the vast majority report being busier than ever before. They admit to sacrificing their relationships, personal time and a good night's sleep for their success. Twenty years ago, 60% of Americans had sit-down dinners with their families, and now only 20% do. But when they're asked what they would do with an eight-day week, they say "Oh that'd be great". They would spend that time labouring away to achieve more. They're constantly trying to get ahead, to get toward a future point of happiness.

So, it's really important to be aware of how other people think about time. We tend to think: "Oh, that person's really irresponsible" or "That guy's power hungry" but often what we're looking at is not fundamental differences of personality, but really just different ways of thinking about time. Seeing these conflicts as differences in time perspective, rather than distinctions of character, can facilitate more effective cooperation between people and get the most out of each person's individual strengths.

Social science questions

Time Perspectives.

Complete the table below. Write **ONE WORD ONLY** for each answer

Time Zone	**Outlook**	**Features & consequences**
Past	Positive	Remember good times

	Negative	Focus on failures.
Present	Hedonistic	Live for Q1] ----------------, seek sensation, avoid pain.
	Fatalistic	Life’s path can’t be changes.
Future	Q2] ----------------	Prefer work to play. Don’t give in to temptation.
	Fatalistic	Have a strong belief in life after death.

Choose the correct letter, **A**, **B** or **C**

Q3] - We are all present hedonists

A- At school
B- At birth
C- While eating and drinking.

Q4] - present-oriented children

A- Do not realize present actions can have negative future effects
B- Are unable to learn lessons from past mistakes
C- Know what could happen if they do something bad, but do it anyway.

Q5] - If Americans had an extra day per week, they would spend it

A- Working harder
B- Building relationships
C- Sharing family meals

5- Lecture: Zoology

Good afternoon everyone. Well, with some of you about to go out on field work it's timely that in this afternoon's session I'll be sharing some ideas about the reasons why groups of whales and dolphins sometimes swim ashore from the sea right onto the beach and, most often, die in what are known as 'mass standings'.

Unfortunately, this type of event is a frequent occurrence in some of the locations that you'll be travelling to, where sometimes the tide goes out suddenly, confusing the animals.However, there are many other theories about the causes of mass strandings.

The first is that the behaviour is linked to parasites. It's often found that stranded animals were infested with large numbers of parasites. For instance, a type of worm is commonly found in the ears of dead whales. Since marine animals rely heavily on their hearing to navigate, this type of infestation has the potential to be very harmful.

Another theory is related to toxins, or poisons. These have also been found to contribute to the death of many marine animals. Many toxins, as I'm sure you're aware, originate from plants, or animals. The whale ingests these toxins in its normal feeding behaviour but whether these poisons directly or indirectly lead to stranding and death, seems to depend upon the toxin involved.

In 1988, for example, fourteen humpback whales examined after stranding along the beaches of Cape Cod were found to have been poisoned after eating tuna that contained saxitoxin, the same toxin that can be fatal in humans.

Alternatively, it has also been suggested that some animals strand accidentally by following their prey ashore in the confusion of the chase. In 1995 David Thurston monitored pilot whales that beached after following squid ashore. However, this idea does not seem to hold true for the majority of mass strandings because examination of the animals' stomach contents reveal that most had not been feeding as thev stranded .
There are also some new theories which link strandings to humans. A growing concern is that loud noises in the ocean cause strandings. Noises such as those caused bv military exercises are of particular concern and have been pinpointed as the cause of some strandings of late.

One of these, a mass stranding of whales in 2000 in the Bahamas coincided closely with experiments using a new submarine detection system. There were several factors that made this stranding stand out as different from previous strandings. This led researchers to look for a new cause. For one, all the stranded animals were healthy. In addition, the animals were spread out along 38 kilometres of coast, whereas it's more common for the animals to be found in a group when mass strandings occur.

A final theory is related to group behaviour, and suggests that sea mammals cannot distinguish between sick and healthy leaders and will follow sick leaders, even to an inevitable death. This is a particularly interesting theory since the whales that are thought to be most social - the toothed whales - are the group that strand the most frequently.

The theory is also supported by evidence from a dolphin stranding in 1994. Examination of the dead animals revealed that apart from the leader, all the others had been healthy at the time of their death.

Without one consistent theory however it is very hard for us to do anything about this phenomenon except to assist animals where and when we can. Stranding networks have been established around the world to aid in rescuing animals and collecting samples from those that could not be helped. I recommend John Connor's Marine Mammals Ashore as an excellent starting point if you're interested in finding out more about these networks, or establishing one yourself.

Zoology questions

Mass Strandings of Whales and Dolphins.

Mass strandings: situations where group of whales, dolphins, etc. swim onto the beach and die

1) - Common in areas where the Q1] --------------------- can change quickly.

Toxins

2) - Poisons from Q2] ---------------- or Q3] ------------------- are commonly consumed by whales. E. g. Cape Cod (1988) - whales were killed by saxitoxin.

Human Activities

3) – Q4] ------------------ from military tests are linked to some recent strandings.

4) - The Bahamas (2000) stranding was unusual because the whales were all Q5] ------ ----- 6- *Lecture: Business*

Good morning. Welcome to this talk on Space Management. And today I'm going to look particularly at space management in the supermarket.
Now since the time supermarkets began, marketing consultants, like us, have been gathering information about customers' shopping habits.

To date, various research methods have been used to help promote the sales of supermarket products. There is, for example, the simple and direct questionnaire which provides information from customers about their views on displays and products and then helps retailers make decisions about what to put where. Another method to help managers understand just how shoppers go around their stores are the hidden television cameras that film us as we shop and monitor our physical movement around the supermarket aisles: where do we start, what do we buy last, what attracts us, etc.

More sophisticated techniques now include video surveillance and such devices as the eye movement recorder. This is a device which shoppers volunteer to wear taped into a headband, and which traces their eye movements as they walk round the shop recording the most eye-catching areas of shelves and aisles. But with today’s technology. Space Management is now a highly sophisticated method of manipulating the way we shop to ensure maximum profit. Supermarkets are able to invest millions of pounds in powerful computers which tell them what sells best and where.

Now, an example of this is Spaceman which is a computer program that helps the retailer to decide which particular product sells best in which part of the store. Now Spaceman works by receiving information from the electronic checkouts (where customers pay) on how well a product is selling in a particular position. Spaceman then suggests the most profitable combination of an article and its position in the store.

So, let’s have a look at what we know about supermarkets and the way people behave when they walk down the aisles and take the articles they think they need from the shelves.

Now here’s a diagram of one supermarket aisle and two rows of shelves. Here’s the entrance at the top left-hand corner.

Now products placed here, at the beginning of aisles, don’t sell well. In tests, secret fixed cameras have filmed shoppers’ movements around a store over a seven-day period. When the film is speeded up, it clearly shows that we walk straight past these areas on our way to the centre of an aisle. Items placed here just don’t attract people.

When we finally stop at the centre of an aisle, we pause and take stock, casting our eyes along the length of it. Now products displayed here sell well and do even better if they are placed at eye level so that the customer’s eyes hit upon them instantly. Products here are snapped up and manufacturers pay a lot for these shelf areas which are known in the trade as hotspots. Naturally everyone wants their products to be in a hotspot.

But the prime positions in the store are the ends of the aisles, otherwise known as Gondola ends. Now these stand out and grab our attention. For this reason new products are launched in these positions and manufacturers are charged widely varying prices for this privileged spot. Also, the end of an aisle may be used for promoting special offers which are frequently found waiting for us as we turn the corner of an aisle.

Well, now, eventually of course, we have to pay. Any spot where a supermarket can be sure we are going to stand still and concentrate for more than a few seconds is good for sales. That’s why the shelves at the checkout have long been a favourite for manufacturers of chocolates — perhaps the most sure-fire “impulse” food of all.

Business questions

Space management

Complete the table, write **NO MORE THAN THREE WORDS** for each answer.

RESEARCH METHOD	INFORMATION PROVIDED
Questionnaire	What costumers think about products
Q1] --------------------------	How costumers move around supermarket aisles.
Eye movement Q2] -----------------------	The most eye-catching areas of the shop

- Label the diagram write **NO MORE THAN THREE WORDS** for each answer.

SUPERMARKET AISLE

ENTRANC -first shelves: costumers usually Q3] ----------------- these.	AISLE Products places at eye level. These areas are known as Q4] -------------------- .	EXIT Checkout. Gondola end: Often find Q5] ------------------------------------ displayed here.

Appendix 2

Participants' consent form

Information page

Using visuals in listening comprehension tests

My name is Iman Elmankush, PhD student. I am currently doing research on the effects of visuals (video and still photos) on students' performance in listening comprehension tests. I am writing to ask if you are able to take part in this study.

Participants in this study will be asked to take a computer based listening comprehension test in the Eye tracker lab at the department of Education. During the visualized test, the participants' eye track will be recorded. A verbal report (interview) will be conducted immediately after completing the test. The whole procedure will take approximately 60 min of your time. **Anonymity**

Any data that you will provide (i.e., test results, eye tracking data, interview records) will be anonymous and securely stored by code number.

Storing and using data

The collected data will be stored in a password protected computer, and it can be used (anonymously) for further analysis or shared for research or training purpose. Please indicate in the enclosed consent form if you happy for this anonymized data to be used in these ways.

You are free to withdraw from the study at any time during data collection.

I hope that you will agree to take part in this study. If, at any time, you have any questions about the study that you would like to ask before giving consent or after the data collection, please feel free to contact the researcher (Iman Elmankush) by email: iye500@york.ac.uk. Or the Chair of Ethics Committee via email: education-research-administrator@york.ac.uk.

If you are happy to participate in the study please complete the enclosed consent form and hand it in to the researcher by 1st of November 2015.

Please keep this information sheet for your own records.

Thank you for taking the time to read this information.

Yours sincerely
Iman Elmankush

Using visuals in listening comprehension tests

Consent form

Please tick each box if you are happy to take part in this research.

I confirm that I have read and understood the information given to me about the □

above named research project and I understand that this will involve me taking part as described above.

I understand that the purpose of the research is to investigate the effects of using visuals on students' performance in listening comprehension tests. □

I understand that data will be stored securely on a password protected computer and only the researcher (Iman Elmankush) will have access to any identifiable data. I understand that my identity will be protected by use of a code. □

I understand that my data will not be identifiable and the data may be used ….

in publications that are mainly read by university academics □

in presentations that are mainly read by university academics □

in publications that are mainly read by the public in presentations that are □

mainly read by the public □

freely available online □

I understand that data could be used for future analysis or other purposes □

I understand that I can withdraw my data at any point during data collection.

□

I understand that I will be given the opportunity to comment on a written record of my responses.

Signature: ____________________

Data: ____ ____ _____

Appendix 3

Pre-test questionnaire

Participant

□

Please answer the following questions

1- Your age: -----------------------

2- Gender: F □ M □

3- Native language(s): -----------------------.

4- Home country: ----------------------------.

5- Your current or expected major at the university of York:

---.

6- Approximate number of years learning English: ------------.

References

Adank, P., Evans, B. G., Stuart-Smith, J., Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 520–529.

Ahuja, P., & Ahuja, G. C. (2007). *How to Listen Better*. Sterling Publishers Pvt. Ltd.

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

Andrychowicz-Trojanowska, A. (2018). Basic terminology of eye-tracking research. *Applied Linguistics Papers*, (25/2), 123-132.

Antes, T. A. (1996). Kinesics: The value of gesture in language and in the language classroom. *Foreign language annals*, *29*(3), 439-448.

Ardila, A., Bernal, B., & Rosselli, M. (2015). Language and visual perception associations: Meta-analytic connectivity modelling of Brodman area 37. *Behavioural neurology*, vol. 2015, Article ID 565871, 14 pages, doi:10.1155/2015/565871.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, 32(1), 25–36.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, 15(9), 578– 582.

Ashby, J., Yang, J., Evans, K. H., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, & Psychophysics*, 74(4), 634-640.

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

Bachman, L. (1995). *An investigation into the comparability of two tests of English as a foreign language* (Vol. 1). Cambridge University Press.

Bachman, L., & Palmer, A. (1983). *Oral interview test of communication proficiency in English.* Los Angeles: Photo-offset.

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

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

Ball, L. J., Eger, N., Stevens, R., & Dodd, J. (2006). Applying the PEEP method in usability testing. *Interfaces*, 67(Summer), 15-19.

Baltova, I. (1994). Impact video on the comprehension skills forced on French students. *Canadian Modern Language Review*, (3), 108-116.

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper*. (TOEFL Monograph Series No. MA-19). Princeton, NJ: Educational Testing Service.

Bloomfield, L., Lane, E., Mangalam, M., & Kelty-Stephen, D. G. (2021). Perceiving and remembering speech depend on multifractal nonlinearity in movements producing and exploring speech. *Journal of The Royal Society Interface*, *18*(181), 20210272.

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. Maryland University College Park.

Bodie, G. D., Janusik, L. A., Välikoski, T. R. (2008). *Priorities of listening research: Four interrelated initiatives*. A white paper sponsored by the research committee of International Listening Association. Retrieved, 2 February 2014, from http://www.listen.org/whitePaper.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. *The on-line study of sentence comprehension: Eye tracking, ERP, and beyond*, 51-76.

Brindley, G. (1998). Assessing listening abilities. *Annual review of applied linguistics*, 18, 171-191.

Brown, H. D. (2007). Teaching by principles: An interactive approach to language pedagogy. White Plains, NY: Longman

Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment Principles and Classroom Practices*. NY: Pearson Education.

Buck, G. (1991). The testing of listening comprehension: an introspective study1. *Language Testing*, 8(1), 67-91.

Buck, G. (1997). The testing of listening in second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment* (pp. 65-74). Dordrecht, The Netherlands: Kluwer.

Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. Language Testing, 15, 119–157

Burgoon, J. (1994). Non-verbal signals. In M. Knapp & G. Miller (Eds.), *Handbook of interpersonal communication* (pp. 344-393). London: Routledge.

Call, M. (1985). Auditory short-term memory, listening comprehension, and the input hypothesis. *TESOL Quarterly*, 19, 765-781.

Camps, J. (2003). Concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks. *International Journal of Applied Linguistics*, 13(2), 201-221.

Canning, J. (2004). *"Disability and residence abroad."* Subject Centre for Languages, Linguistics and Area Studies Good Practice Guide. Retrieved 7 February 2017, from http://www.llas.ac.uk/resources/gpg/2241.

Carpenter, P., & Just, M. (1983). What your eyes do while your mind is reading-In K. Rayner (Ed.), *Eye movements in reading: Perceptual and Language Processes* (pp. 275-307). New York: Academic Press.

Carrell, P. L., Dunkel, P. A., & Mollaun, P. (2002). *The effects of notetaking, lecture length and topic on the listening component of the TOEFL 2000.* (TOEFL Monograph Series No. MS-23). Princeton, NJ: Educational Testing Service.

Carroll, J. B. (1983). Psychometric theory and language testing. In J. Oller (Ed.), *Issues in language testing research* (pp. 80-107). New York: W. H. Freeman.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62.

Castet, E., Jeanjean, S., & Masson, G. S. (2002). Motion perception of saccade-induced retinal translation. *Proceedings of the National Academy of Sciences*, *99*(23), 15159-15163.

Celce-Murcia, M. (2002). *Teaching English as a second or foreign language* (3rd ed.). U.S.A: Heinle & Heinle Publishers. Rd.

Chang, A. C., & Read, J. (2008). Reducing listening text anxiety through various forms of listening support. *TESL-EJ*, 12(1), 1–25.

Chapelle, C. A. (1998). JL Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.) *Interfaces between second language acquisition and language testing research* (32-70). Cambridge University Press.

Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL quarterly*, 26(2), 345374.

Chung, U. K. (1994). *The effect of audio, a single picture, multiple pictures, or video on second-language listening comprehension* (Doctoral dissertation), University of Illinois at Urbana-Champaign.

Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149-210.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I., (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34, 696–702.

Coniam, D. (2002). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System* 29, 1–14.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research (*2nd ed.). Los Angeles; London: SAGE.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, 10(1), 1-8.

Derwing, T., & Munro, M. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics*, 22(3), 324–337.

Douglas, D., & Hegelheimer, V. (2007). Assessing language using computer technology. *Annual Review of Applied Linguistics*, 27, 115-132.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer Science & Business Media.

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly*, 25, 431-457.

Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it.* 1 (pp. 119-128). British Computer Society.

Elliott, M. (2013). Test taker characteristics. In Geranpayeh, A & Taylor, L (Eds.), *Examining Listening: Research and Practice in Assessing Second Language Listening, Studies in Language Testing*, 35, (pp. 36- 76). Cambridge University Press.

Elliott, M., & Wilson, J. (2013). Context validity. In Geranpayeh, A & Taylor, L (Eds.), *Examining Listening: Research and Practice in Assessing Second Language Listening, Studies in Language Testing*, 35, (pp. 152-241). Cambridge University Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1), 19-23.

Farhady, H. (1983). New directions for ESL proficiency testing. *Issues in language testing research*, 253-269.

Field, J. (2004). *Psycholinguistics: The key concepts*. Psychology Press.

Field, J. (2008). Guest editor's introduction emergent and divergent: A view of second language listening research. *System*, 36(1), 2-9.

Field, J. (2009). A cognitive validation of the lecture-listening component of the IELTS listening paper. *IELTS research reports*, 9, 17-66.

Field, J. (2010). Listening in the language classroom. *ELT journal*, 64(3), 331-333.

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes*, 10(2), 102-112.

Field, J. (2012). 1 The cognitive validity of the lecture-based question in the IELTS Listening paper. *IELTS Collected Papers 2: Research in Reading and Listening Assessment,* 2, 391.

Field, J. (2013). Cognitive validity. *Examining listening*, 77-151.

Flowerdew, J. (1994). *Academic listening: Research perspectives*. Cambridge University Press.

Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16(1), 27–46.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge University Press.

Flowerdew, J., & Miller, L. (2010). Listening in a second language. *Listening and human communication in the 21st century*, 158-177.

Fortune, A. (2004). *Testing listening comprehension in a foreign language - does the number of times a text is heard affect performance*? Unpublished M.A. dissertation.

Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's mini-talks. *Language Testing,* 16(1), 2–32.

Gallagher, C. J. (2003). Reconciling a tradition of testing with a new learning paradigm. *Educational Psychology Review,* 15(1), 83-99.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly,* 10(1), 9-27.

George, D. & Mallery, M. (2010). *Using SPSS for Windows step by step: a simple guide and reference*. Boston, MA: Allyn & Bacon.

Geranpayeh, A., & Taylor, L. (2008). Examining Listening: developments and issues in assessing second language listening. *Research Notes* 32, 2-5.

Geranpayeh, A., & Taylor, L. (2013). *Examining Listening: Research and practice in assessing second language listening* (Vol. 35). Cambridge University Press.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40, 97–118.

Ginther, A (2002) Context and content visuals and performance on listening comprehension stimuli, *Language Testing* 19 (2), 133–167.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.

Goh, C. C., & Hu, G. (2013). Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language Awareness*, 23(3), 255-274. Retrieved 15. 3. 2016. http://dx.doi.org/10.1080/09658416.2013.769558.

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in timebenefit analysis. *Language Learning*, 40(3), 311–336.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly*, 26(2), 385– 390.

Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal*, 15, 85-88.

Gruba, P. (1997). The role of video media in listening assessment. *System*, 25, 335-345.

Gruba, P. (2006). Playing the videotext: A media literacy perspective on video-mediated L2 listening. *Language Learning & Technology*, 10(2), 77-92.

Gu, Y. (2018). Two-Way Listening. *The TESOL Encyclopedia of English Language Teaching*, 1-8.

Hale, G. A., & Courtney, R. (1994). The effects of note-taking on listening comprehension in the Test of English as a Foreign Language. *Language Testing*, 11(1), 29–47.

Halliday, M. A., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective* (2nd ed.). Oxford and New York: Oxford University Press.

Harding, L. (2008). Accent and academic listening assessment: A study of test-taker perceptions. *Melbourne Papers in Language Testing*, 13(1), 1-3.

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy*, 6(2), 267-274.

Henderson, J. M., Olejarczyk, J., Luke, S. G., & Schmidt, J. (2014). Eye movement control during scene viewing: Immediate degradation and enhancement effects of spatial frequency filtering. *Visual Cognition*, *22*(3-4), 486-502.

Henning, G. (1990). *A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance*. (TOEFL Research Reports RR-33). Princeton, NJ: Educational Testing Service.

Hillier, Y. (2002). *Reflective teaching in further and adult education*. A&C Black.

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PloS one*, 10(8), e0136905.

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

Hu, G. G., & Jiang, N. (2011). Semantic integration in listening comprehension in a second language: Evidence from cross-model priming. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research: insights from psycholinguistics* (pp.199-218). Amsterdam, Netherlands: John Benjamins Publishing Company.

Hughes, A. (2003). *Testing for language teachers*. Cambridge university press .

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, 6(2), 231-255.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In Hyona J, Radach R, Deubel H (Eds.) *The Mind's eye: cognitive and applied aspects of eye movement* (pp. 573605). Oxford.

James, W. (1981). *The Principles of psychology* (Vol. 1) Cambridge, MA: Harvard University Press.

Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228-242.

Jeon, J. (2007). *A study of listening comprehension of academic lectures within the construction-integration model* (Doctoral dissertation). The Ohio State University.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Lawrence Erlbaum Associates, Inc.

Jones, G., Pearson, L., & Glyn, U. K. (2011). Research Summary: Once or twice? A critical review of current literature on the question how many times the audio recording should be played in listening comprehension testing items. *Listening Comprehension Testing Items*.

Jordanova, L. (2016). Approaching visual materials. *Research methods for history*, 30-47.

Just, M. A., & Carpenter, P. A. (1984). A Theory of reading: From Eye fixations to comprehension. *Psychological Review,* 87(4), 329–354.

Kelly, P. (1991). Lexical Ignorance: The Main Obstacle to Listening Comprehension with Advanced Foreign Language Learners. *Iral*, 29(2), 135-49.

Kienbaum, B. E., & Russell, A. J. S. Welty. (1986). *Communicative Competence in Foreign Language Learning with Authentic Materials*.

King, P. (1994). 1 1 Visual and verbal messages in the engineering lecture: note taking by postgraduate L2 students. *Academic listening: Research perspectives*, 219.

King, P. E., & Behnke, R. R. (1989). The effect of time-compressed speech on comprehensive, interpretive, and short-term listening. *Human Communication Research*, 15(3), 428-443.

Kostelnick, C., & Roberts, D. D. (2011). *Designing visual language: Strategies for professional communicators*. Longman.

Kostin, I. (2004). *Exploring item characteristics that are related to the difficulty of TOEFL dialogue items*. (TOEFL Research Report RR-79). Princeton, NJ: Educational Testing Service.

Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests. A Teacher's Book*. New York: McGraw-Hill.

Li, Z. (2013). The issues of construct definition and assessment authenticity in videobased listening comprehension tests: Using an argument-based validation approach. *International Journal of Language Studies*, 7(2), 61-82.

Lin, M. (2006). *The effects of note-taking, memory and rate of presentation on EFL learners' listening comprehension*. Unpublished doctoral dissertation, La Sierra University, California.

Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics*, 17(1).

Long, D. R. (1989), Second Language Listening Comprehension: A Schema-Theoretic Perspective. *The Modern Language Journal*, 73, 32–40.

Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The modern language journal*, 75(2), 196-204.

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics*, 18, 3- 19.

Lynch, T. (2009). *Teaching second language listening*. Oxford University Press.

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10(2), 79-88.

Macaro, E., Vanderplank, R., & Graham, S. (2005). *A systematic review of the role of prior knowledge in unidirectional listening comprehension*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Major, R., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2005). Testing the effects of regional, ethnic and international dialects of English on listening comprehension. *Language Learning*, 55, 37–69.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational psychologist*, 32(1), 1-19.

Mayer, R. E. (2005). Cognitive theory in multimedia learning. In R. E. Mayer (Ed). *The Cambridge handbook of multimedia learning* (pp.31-48). Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63(8), 760.

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge, UK: Cambridge University Press

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381-401.

McIntyre, N. (2016). *Teach at first sight: Expert teacher gaze across two cultural settings* (Unpublished Doctoral thesis). University of York. UK.

Messick, S. (1996). Validity and washback in language testing. *ETS Research Report Series*, 1996(1).

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013). Visual information can hinder working memory processing of speech. *Journal of Speech, Language, and Hearing Research*, 56(4), 1120-1132.

Moore, D. M., & Dwyer, F. M. (Eds.). (1994). *Visual literacy: A spectrum of visual learning*. Educational Technology.

Morley, J. (2001). Aural comprehension instruction: Principles and practices. In M. CelceMurcia (Ed.), *Teaching English as a second or foreign language* (pp. 69-85). Boston, MA: Heinle & Heinle.

Mueller, G. A. (1980). Visual contextual cues and listening comprehension: An experiment. *The Modern Language Journal*, 64(3), 335-340.

Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. (ETS Research Report 95-37). Princeton, NJ: Educational Testing Service.

Noton, D., & Stark, L. (1971). Scan paths in saccadic eye movements while viewing and recognizing patterns. *Vision research*, 11(9), 929-942.

Ockey, G. J. (2007). Construct implications of including still image or video in computerbased listening tests. *Language Testing*, 24(4), 517-537.

Ockey, G. J. (2009). Developments and Challenges in the Use of Computer-Based Testing for Assessing Second Language Ability. *The Modern Language Journal*, 93(s1), 836-847.

Ockey, G. J., & French, R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, *37*(5), 693-715.

Ockey, G. J., & Wagner, E. (2018). *Assessing L2 listening: Moving towards authenticity* (Vol. 50). John Benjamins Publishing Company.

Oller, J. (1979). *Language tests at school*. London: Longman.

Oller, J. (1983). A consensus for the eighties? In J. Oller (Ed.), *Issues in language testing research* (pp. 351-356). Rowley, MA: Newbury House.

Olson, K. (2003). LSAT Listening Assessment: Theoretical Background and Specifications. *Law School Admission Council (LSAC) Research Report 03–02.* Retrieved from http://www.lsac.org/lsacresources/Research/rr/pdf/RR-03-02.pdf.

Osada, N. (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue*, 3(1), 53-66.

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

Parkhurst, D., & Niebur, E. (2003). Scene Content selected by active vision. *Spatial Vision,* 16(2), 125–154.

Pateşan, M., Balagiu, A., & Alibec, C. (2018). Visual aids in language education. In International Conference Knowledge-Based Organization (Vol. 24, No. 2, pp. 356-361).

Peterson, P. W. (2001). Skills and strategies for proficient listening. *Teaching English as a second or foreign language*, 3, 87-100.

Pettersson, R. (2002). *Information design: An introduction* (Vol. 3). John Benjamins Publishing.

Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. In Fincher S., Markopoulos P., Moore D., Ruddle R. (Eds.) *People and computers XVIII—Design for life* (pp. 363378). Springer. London.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2), 160.

Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal*, 14(1), 34-44.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European journal of cognitive psychology*, *16*(1-2), 3-26.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology*, 62(8), 1457-1506.

Reingold, E. M. (2014). Eye tracking research and technology: Towards objective measurement of data quality. *Visual cognition*, 22(3-4), 635-652.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL quarterly*, 17(2), 219-240.

Rolfs, M. (2015). Attention in active vision: A perspective on perceptual continuity across saccades. *Perception*, *44*(8-9), 900-919.

Rosenhouse, J., Haik, L., & Kishon-Rabin, L. (2006). Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals. *International Journal of Bilingualism*, 10(2), 119–135.

Rost, M. (2006). Areas of research that influence L2 listening instruction. In E. Uso-Juan & A. Martinez-Flor (Eds.) *Current Trends in the Development and Teaching of the Four Language Skills* (pp. 47–74). New York: Mouton de Gruyter.

Rost, M. (2013). *Teaching and researching: Listening*. Routledge.

Rost, M. (2014). Listening in a multilingual world: The challenges of second language (L2) listening. *International Journal of Listening*, 28(3), 131-148.

Rost, M., & Candlin, C. N. (2014). *Listening in language learning*. Routledge.

Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers A genre-based study. *English for specific purposes*, 21(1), 19-40.

Royce, T. D. (2007). Multimodal communicative competence in second language contexts. In T. D. Royce & W. L. Bowcher (Eds.) *New directions in the analysis of multimodal discourse* (pp. 361-390). Psychology Press.

Rubin, J. (1990). Improving foreign language listening comprehension. *Georgetown University round table on languages and linguistics*, Georgetown University press. 309-316.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199–221.

Rubin, J. (1995). The contribution of videos to the development of competence in listening. In D. J. Mendelsohn & J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 151- 165). San Diego, CA: Dominie Press.

Rupp, A. A., Garcia, P, & Jamieson, J. (2001). Combining multiple regression and CART to understand difficulty in second language reading and listening comprehension test items. *International Journal of Testing*, 1(3 & 4), 185–216.

Safarali, S. & Hamidi, H. (2012). The impact of videos presenting speaker's gestures and facial clues on Iranian EFL learners' listening comprehension. *International Journal of applied Linguistics & English Literature*, 1(6), 106-114.

Sasaki, T. (2003). Recipient orinentation in verbal report protocols: Methodological issues in concurrent think-alaud. *Second Langauge Studies*, 22 (1), 1-54.

Scheiter, K., & Van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, 23(9), 1209-1214.

Schnapp, D. C. (1991). The Effects of Channel on Assigning Meaning in the Listening Process. *International Journal of Listening*, *5*(1), 93-107.

Schnotz, W. (2002). Commentary: Towards an integrated view of learning from text and visual displays. *Educational psychology review*, 14(1), 101-120.

Schnotz, W. (2005). An integrated model of text and picture comprehension. *The Cambridge handbook of multimedia learning*, 49-69.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and instruction*, 13(2), 141-156.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. *The science of real-time data capture: Self-reports in health research*, 11-26.

Secules, T., Herron, C., & Tomasello, M. (1992). The effect of video context on foreign language learning. *The Modern Language Journal*, 76(4), 480-490.

Shepherd, J. (2012). *University students spend no more time with lecturers than six years ago*. Higher Education Policy Institute (HEPI) Report, London: HEPI.

Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language testing*, 14(2), 185-213.

Shriver, K. A. (1997). *Dynamics in document design: Creating text for readers*. New York, NY: John Wiley & Sons.

Siemer, M., & Reisenzein, R. (1998). Effects of mood on evaluative judgements: Influence of reduced processing capacity and mood salience. *Cognition and Emotion,* 12(6), 783–805.

Smidt, E. & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning*, 17 (5), 517-556.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience*, 32(25), 8443-8453.

Spolsky, B. (1968). Language testing: the problem of validation. *TESOL Quarterly*, 2(2), 88-94.

Spolsky, B. (1968). What Does it Mean to Know a Language, Or How Do You Get Someone to Perform His Competence? In J. W. Oller & Richards (Eds.), *Focus on the learner* (pp. 164-176). Rowley, Mass: Newbury House.

Sueyoshi, A. & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning,* 661-699.

Suvorov, R. (2009). Context visuals in L2 listening tests: The effects of photographs and video vs. audio-only format. *Developing and evaluating language learning materials*, 53-68.

Suvorov, R. (2011). The effects of context visuals on L2 listening comprehension. *Editorial notes*, 2.

Suvorov, R. (2013). *Interacting with visuals in L2 listening tests: An eye-tracking study.* Unpublished doctoral dissertation. Iowa State University. Ames, Iowa.

Suvorov, R., & Hegelheimer, V. (2013). Computer-Assisted Language Testing. In A. J. Kunnan (Ed.), *The Companion to Language Assessment* (pp. 593–613). Malden, MA: Wiley-Blackwell.

Tayler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, 51(2), 257–280.

Taylor, L. (2013). Introduction. In Geranpayeh, A & Taylor, L (Eds.), *Examining Listening*. *Research and Practice in Assessing Second Language Listening, Studies in Language Testing*, 35 (pp. 1- 35). Cambridge University Press.

Taylor, K. L., & Dionne, J. P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, 92(3), 413.

Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied linguistics*, 19(4), 432-451.

Van Gog, T., Paas, F., van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237.

Vandergrift, L. (2004). 1. Listening to Learn or Learning to Listen? *Annual Review of Applied Linguistics*, 24, 3-25.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(3), 191–210.

Vandergrift, L. (2015). Researching listening. *Research Methods in Applied Linguistics: A Practical Resource*, 299.

Vandergrift, L., & Goh, C. C. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, 60(2), 470-497.

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, *22*(3), 293-301.

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual, and motor control aspects. In Kowler, E. (Ed.), *Reviews of oculomotor research: Eye Movements and Their Role in Visual and Cognitive Processes* (pp. 353-393). Amsterdam: Elsevier Science

Vollmer, H., & Sang, F. (1983). Competing hypotheses about second language ability: A plea for caution. In J. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 29-79). Rowley, MA: Newbury House.

Von Raffler-Engel, W. (1980). Kinesics and Paralinguistics: A Neglected Factor in Second-Language Research and Teaching. *Canadian Modern Language Review*, 36(2), 225-37.

Wagner, E. (2002). *Video listening tests: A pilot study.* Teachers College Columbia University Working Papers in TESOL and Applied Linguistics, 2, 1.

Wagner, E. (2006a). Can the search for "fairness" be taken too far? *TESOL & Applied Linguistics.* Vol. 6, No. 2. Columbia University.

Wagner, E. (2006b). *Utilizing the visual channel: An investigation of the use of videotexts on tests of second language listening ability* (Unpublished doctoral dissertation). Teachers College, Columbia University, New York.

Wagner, E. (2007). Are they watching? Test-takers viewing behavior during an L2 video listening test. *Language learning & technology*. Vol. 11, P 67-86.

Wagner, E. (2008). Video listening tests: what are they measuring? *Language Assessment Quarterly*, *5*(3), 218-243

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language testing*.

Walker, N. (2014). Listening: The most difficult skill to teach. *Encuentro, 23*(1), 167-175.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication,* 50, 81–94.

Watts, K. (2001). Focus on Register in the Spanish Language Classroom. ERIC

Weir, C. J. (1990). *Communicative language testing*. UK: Prentice Hall.

Weir, C. J. (1993). *Understanding and developing language tests*. UK: Prentice-Hall.

Whyte, J., Cormier, E., & Pickett-Hauber, R. (2010). Cognitions associated with nurse performance: a comparison of concurrent and retrospective verbal reports of nurse performance in a simulated task environment. *International journal of nursing studies,* 47(4), 446-451.

Wiswede, D., Rüsseler, J., & Münte, T. F. (2007). Serial position effects in free memory recall: An ERP study. *Biological Psychology,* 75, 185–193.

Witkin, B. R., & Trochim, W. W. (1997). Toward a synthesis of listening constructs: A concept map analysis. *International Journal of Listening,* 11(1), 69-87.

Wolfersberger, M. (2013). Refining the construct of classroom-based writing-fromreadings assessment: The role of task representation. *Language Assessment Quarterly,* 10(1), 49-72.

Wolvin, A., & Coakley, C. G. (1996). *Listening* (5th ed.). Madison, WI: Brown & Benchmark.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum press.

Yeldham, M., & Gruba, P. (2014). Toward an instructional approach to developing interactive second language listening. *Language Teaching Research*, *18*(1), 33-53.

Y'ian, W. (1998). What do tests of listening comprehension test? - A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing* 15(1), 2144.

Ying-hui, H. (2006). An investigation into the task features affecting EFL listening comprehension test performance. *The Asian EFL Journal Quarterly,* 8(2), 33–54.

Zwaan, R. A., Kaup, B., Stanfield, R. A., & Madden, C. J. (2001). Language comprehension as guided experience. (http://cogprints.soton.ac.uk/documents/).

Zhou, G. L., & Yang, S. D. (2004). The effects of visual aids on English major's listening comprehension. *Journal of PLA Foreign Languages Institute,* (3).

Zielinski, S. (2006). *Deep time of the media: Towards an archaeology of hearing and seeing by technical means.* Cambridge, MA: MIT Press.