The effects of applying multimodality in oral comprehension tasks in the English classroom of Batxillerat: students' performance and attitudes

Treball de Final de Màster Universitari en Formació del Professorat d'Educació Secundària Obligatòria i Batxillerat, Formació Professional i Ensenyaments d'Idiomes (especialitat llengua anglesa)

Author: Maria del Mar Sagué Guillén

Academic year: 2020-2021

Supervisor: Dr. M. Àngels Pinyana

Vic, June 2021

ACKNOWLEDGEMENTS

First, I would like to thank my tutor, Dr. Àngels Pinyana, for her constant feedback, support, patience and time in the development of this research, and for motivating me to do my best. This TFM thesis would have not been possible without her assistance. Also in the university field, I would like to thank Dr. Llorenç Comajoan to have put faith in my research since the decision phase, encouraging me to continue with this object of study.

Also, I would like to express my gratitude to Institut Rafael Campalans in Anglès, the school where I was brought up, and its students to have allowed the methodology to be carried out throughout my Practicum in the school. Without the sample provided, this research would not exist.

Last but not least, I also feel very grateful of my family and friends, for their emotional support in the process of this thesis and, more generally, in the intense school year of this Master's degree that will allow me to start with my teaching career very soon.

RESUM

La comprensió oral és una habilitat important quan aprenem una llengua estrangera. Amb l'evolució de la tecnologia actualment, les tasques de *listening* poden incloure més que el tradicional input només auditiu, com poden ser imatge i text escrit. La combinació de múltiples modes és coneguda com a multimodalitat. Aquest estudi pretén descobrir si la comprensió augmenta quan apliquem la multimodalitat (escoltar i visualitzar) i quines són les actituds i percepcions dels aprenents. 22 participants –estudiants de Batxillerat– han dut a terme dues tasques de *listening* (només escolta –LO–, i escolta i visualització – LW) i han omplert un qüestionari. Per una banda, la comprensió s'ha mesurat quantitativament segons els resultats en la tasca LO i la tasca LW. Per l'altra, s'han examinat les actituds i preferència a través de diferents paràmetres, tant quantitativament com qualitativament, per saber-ne la percepció. Els resultats indiquen que la comprensió augmenta en les tasques de LW si ho comparem amb les de LO i, doncs, que la multimodalitat ajuda a la comprensió. Pel que fa a les actituds, la percepció i la preferència, els participants atribueixen reiteradament característiques a les tasques de LW, com l'entreteniment, la retenció de la informació a la memòria, una major comprensió de la informació general, entre altres. També, afirmen que prefereixen el mode LW unànimement i un alt percentatge creu que és una manera de millorar les seves habilitats en la llengua.

Paraules clau: comprensió oral en anglès com a llengua estrangera, multimodalitat, comprensió oral i visualització, comprensió oral, actituds vers la multimodalitat.

ABSTRACT

Listening is an important skill when learning a foreign language. With the evolution of technology in current times, listening comprehension tasks can include more than the traditional audio-only input, such as image and written text. The combination of multiple modes is known as multimodality. This study aims to discover whether comprehension improves when multimodality (listening and watching) is applied and what the learners' attitudes and perception are. 22 participants –students of Batxillerat– have carried out two listening tasks (listening only –LO–, and listening and watching –LW) and completed a questionnaire. On the one hand, comprehension has been measured quantitatively in performance in LO and LW tasks. On the other hand, both quantitatively and qualitatively, mode attitudes and preference through several parameters have been examined to know students' perception. The findings show that comprehension is enhanced in listening and watching tasks compared to listening only tasks, and, thus, that multimodality helps in comprehension. Concerning attitudes, perception and preference,

participants recurrently attribute characteristics to LW, such as entertainment, information retrieval and easier general information understanding, among others. Also, they unanimously claim to prefer the LW mode and a high percentage perceive it as a way to improve their language skills.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1   INTRODUCTION

1.1   Topic and object of study

It is known that the listening skill might be one of the most difficult abilities when learning a foreign language. Textbooks and many ESL materials, even the newest and less traditional ones, offer oral comprehension materials most of the time in audio format. Our era has achieved a big technological breakthrough, which means that much of what we perceive is in a multimodal way (voice, music, still and moving images, integrated text). As a matter of fact, we know that, when we do not know anything in a foreign language, having visual aids helps us understand a communicative situation that we might not be able to understand with only listening to voice. Multimodality is a widely spread concept for making meaning through different modes, as Legros & Crinon (2002) state: 'multimodality makes sensory information accessible in diverse semiotic codes and offers the opportunity to comprehend information through different channels'. In multimodality, thus, some clues might show up, especially if what is explicitly said appears in the image. Therefore, we do not only understand through our ears. Moreover, multimodal components may seem more visually appealing, specially to the individuals born in the technological era. Hance, why do not we apply these methodologies to the materials in the classroom?

Based on the assertion which numerous studies have already made that oral comprehension increases if multimodal elements appear, this investigation aims to analyse which differences appear in comprehension, concerning task results, and, also, to observe if the students' attitudes and perception are positive towards this methodology, closer to the current reality. In the framework of a Master's in teaching, the sample taken for the study are Secondary education students.

1.1.1   Objectives

As a guideline to orientate my investigation, this study proposes the following objectives:
1. To know the differences in comprehension between audio-only and multimodal formats, while analysing whether oral comprehension scores differ between materials that consist of only voice (LO – listening only) and materials which include multiple modes (LW – listening and watching).
2. To verify that the students' comprehension in the tasks that include multimodal elements (LW) is higher, compared to those that only consist of aural information (LO).
3. To analyse the students' attitudes and perception towards LO and LW tasks and to determine whether tasks in one mode or another increase students' motivation.

1.2 Justification of the interest and originality

I must admit that the listening skill has never been my strength as a learner, and I attribute this to the fact that it is the ability which, under the context of an exam in ESL teaching, can cause more anxiety to learners. According to some authors, the reason for this is that listening is a complex process, consisting of three phases: audition, perception and comprehension (Marrero, 2001, cited in Pérez Bernabeu, 2019). It is a process which goes from the mere reception to the access to meanings.

When looking for new methodologies for enhancing students' listening skills in real contexts, from my standpoint, this investigation in a classroom scale becomes appropriate and interesting. This is a study spurred by the criticism in application of methodologies –and more specific, materials, which do not match the current needs that learners may have in the present day. In the Secondary education classroom it is where we can find the well-known 'digital natives' –term first coined by Prensky (2001, p. 1), who have grown up in a context that does not seem to conceive listening in isolation in a multimedia world, if it is not the case of channels whose nature is unimodal, such as the radio, podcasts or music. Thus, for this reason the present investigation is original, because it aims to see the effect of applying materials closer to nowadays reality to the instruction of foreign language (FL).

2 THEORETICAL BACKGROUND

2.1 Multimodality in language teaching

2.1.1 What is multimodality?

The term multimodality was coined in the mid-19th century and it is focused on disciplines such as 'linguistics, media studies, new literacy studies, education, sociology and psychology' (Jewitt et al., 2016, p. 1). As it can be inferred from its etymology, *multimodality* is the whole of multiple communication 'modes', or, as some other authors prefer to call it, 'sources'. In short, the several theories in multimodality (Kress & Van Leuuwen, 2001 & 2006; Jewitt, 2009; Bezemer & Jewitt, 2010, among others) have shown how different communication modes, such as language, pictures, graphics, sounds, music, gestures, etc., build meaning, For instance, if we find a reading text without pictures, it is a single mode, while if we add some images, music or speaking, it involves multiple modes.

According to Martínez Lirola (2013), 'these modes are semiotic resources that allow meaning to be created through one of them independently or through the combination of many of them simultaneously' (p. 4). This author also states that 'image plays an essential role in multimodal communication due to its capacity to grab attention' (Kress & Van Leewen, 2006; Kress, 2010; Unsworth, 2010, cited in Martínez Lirola, 2013, p. 4).

2.1.2 Multimodality in the technological era

The interest and presence of multimodality has increased as technology has also evolved concurrently. Since the birth of the so called 'digital natives' (Prensky, 2001, p. 1) and the revolution of the Internet in the 1900s, society has become familiar with technological gadgets, such as smartphones, computers and video cameras. Hence, this has led to access to music, images and videos easily. This seems so innate in the digital natives that 'its acquisition has been compared to that of their mother tongue' (Norte Fernández-Pacheco, 2016, p. 51). Even if now multimodal contexts take place constantly, as a result of the change in the way we communicate with each other, in the past, when technology did not still exist, some modes were already combined as we still do nowadays, such as written text and pictures (Bezemer & Kress, 2014). An example of this are comics or illustrated books. In the present, moving images, still images, typography, sounds and colours collaborate as semiotic sources and are thought to facilitate comprehension (Norte Fernández-Pacheco, 2016).

2.1.3 Multimodality in the EFL classroom

Even if it is now that technology has fostered multimodal practices, in the language teaching area, many teachers have already used multimodality in an intuitive way, for example, by mixing image together with a text, as Knox (2008) states:

> Multimodal perspectives on language and language education have only recently appeared on L2 teaching and learning. A brief consideration of the classroom practices of teachers and students shows very quickly, though, that multimodality is something that language teachers have understood intuitively for a long time (p. 140).

In fact, multimodality is a field that has been widely researched when developing the different skills in EFL learning. On the one hand, there have been studies seeking how aural input can improve reading comprehension skills, that is, reading through listening, such as Chang & Millett (2015). Chang (2009) concludes that 'aural-written verification helps L2 learners develop auditory discrimination skills, refine word recognition and gain awareness of form-meaning relationships' in the long term. The study also supports features such as better task performance, perception of shorter duration, and more motivation and attention (p. 652). Likewise, Diao et al. (2007) support the inclusion of written material with spoken material, on the basis that listening through reading facilitates comprehension and recall of information (p. 251). Kartal & Simsek (2017), for their part, also reach the same conclusion with the use of audiobooks in EFL listening comprehension. Nevertheless, Diao et al. (2007) point to the elimination of visual text when the aim of instruction is teaching to listen and not enhance overall comprehension (p. 251).

There is also research in the effects of video captions on vocabulary learning and listening comprehension. Hsieh (2019) states that 'caption type [do] affect vocabulary learning' and '[facilitate] word form and meaning'. Other investigations also focus on image as means to foster vocabulary, comparing the methods of keyword and pictorial, concluding that the keyword method (based on mnemonic acquisition) offers mental imagery, both verbal but also visual (Tavakoli & Gerami, 2012). The keyword method is a mnemonic technique that enhances vocabulary retention through acoustic or orthographic similarities with the L2 word, creating, therefore, a strong association between the target word and the keyword built to increase its memorability (Hulstijn, 1997, p. 204, cited in Tavakoli & Gerami, 2012).

All this recent research comes to the conclusion that the traditional mode that each skill provides can increase learners' performances when being combined. In fact, communication is carried out through 5 different modes, making it multimodal and asserting that communication is not only linguistic or aural information (see Figure 1). Related to the 4 traditional skills, thus, there has recently been a new conception of skills, including 'viewing' and 'visually representing' as new skills in the new era, and some countries have already collected them in language curricula. In Donaghy's works, he refers to viewing as the 'fifth' skill and is defined as 'reading –analysing, evaluation and appreciating– visual texts' (2019), which reinforces the idea that communication is multimodal indeed.
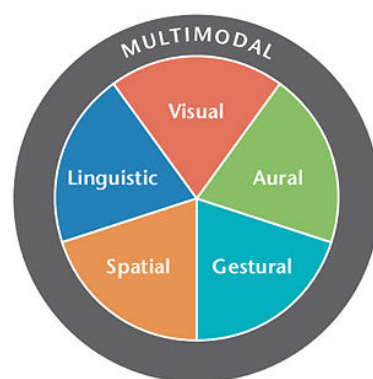
Figure 1. The 5 modes of communication.
Source: the New London Group

2.2 The listening skill

2.2.1 Definition and process of listening

When learning a new language, the aural material seems to be the first material which we face to understand the communicative elements of that language. For its unique features, which include air channel communication, continuums, coarticulation and spontaneity, the listening skill becomes a challenge for the learner (Pérez Bernabeu, 2019, p. 48).

There have been many definitions of listening in FL learning, some of which suggest similar patterns as the following: 'Listening is an active, purposeful processing of making sense of what we hear' (Helgesen, 2003, p. 14) or '(Listening is the) mental process of constructing meaning from spoken input' (Rost, 2002, p. 279). In both cases, the elements of 'making sense' and 'hearing/spoken' appear. Thus, listening means making sense –sounds and words, grammatical patterns and knowledge of the topic– of what you hear –that is, aural input. Also remarkable is that most of the recent definitions include the notion of 'an active skill', as we see a clear difference with the process of 'hearing', which is 'passive'. In any case, it is evident

that it consists of a receptive skill. Some other authors distance the notion of only aural information that the other definitions conveyed and describe listening with aural together with visual information: '(Listening is the action of) processing information which listeners get from visual and auditory clues in order to define what is going on and what the speakers are trying to express' (Rubin, 1995, p. 151). From this proposal of definition, we can notice listening from auditory input as well as with an influence of audio-visual cues.

Listening is a complex process and requires familiarisation with the meaning of words to be effective. In fact, in order to make listening communication efficient, the listener needs to be familiar, and able to recognise, about 90% of the content words (Waring & Nation, 1997). However, listening implies much more than only word recognition. Next, several elements that have great importance when understanding the process of listening are presented, eliciting the complexity of this ability.

a) Purpose

A characteristic of the listening skill is that it can either be with a transactional or interactional purpose. These two communicative purposes differ from each other. On the one hand, interactional –or two-way– use has a social orientation and has the aim to satisfy participants' needs, with interaction (for instance, conversations), while, on the other hand, transactional – or one-way– purposes are more message-orientated and there is not a chance for clarification (Vandergrift, 2011).

b) Processing

In order to conceive listening, we also need to think about the way human beings process received inputs. To start with, processing can be either bottom-up or top-down (Flowerdew & Miller, 2005, p. 24-27). On the one hand, listeners follow a top-down process when they make use of background knowledge to make meaning of the received message. Prior knowledge includes content knowledge, coming from experience and prior knowledge of the topic, and textual knowledge, that is, knowledge about language, the context, the culture in which it takes place (Flowerdew & Miller, 2005). On the other hand, listeners follow a bottom-up processing if meaning making comes from the strict language knowledge, starting from the lower level and 'comes up' to the highest levels to get to the final message: from sounds, to words, to grammatical relationships until lexical meanings (Flowerdew & Miller, 2005, cited in Vandergrift, 2011). While top-down processing might imply getting general meanings, as it happens with extensive reading (also called 'listening for gist'), bottom-up processing alludes to listening for details and specific information. Listeners do not process in only one of the processes, but they rather combine previous knowledge and specific linguistic knowledge.

Directly related to this, it is also relevant to look at the concepts indicated in the image below: intelligibility and comprehensibility. Intelligibility stands for the recognition of

individual words –closer to bottom-up, whereas comprehensibility goes one step further, implying the full understanding of the message, which can be possible through the use of linguistic context –closer to top-down– even if some words might have been missed in intelligibility (Smith & Nelson, 1985, cited in Zoghbor, 2016). As it happened before, it consists of two elements that interact and both somehow help comprehension.
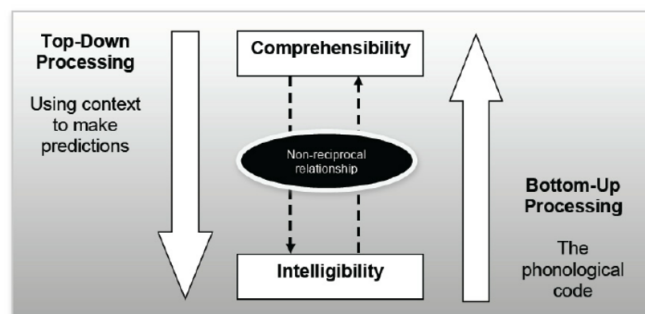
Figure 2. Proposed relationship between comprehensibility and intelligibility and top-down and bottom-up processing. Source: Zoghbor, Z. (2016)

c) Phases

Human hearing is produced in three successive phases: (1) audition, (2) perception, and (3) comprehension (Marrero, 2001, p. 6, cited in Pérez Bernabeu, 2019). The first phase is the mere reception of a sound. Regarding the two following phases, perception has to do with phonetics and how we perceive the sounds, whereas comprehension is a step forward and implies the access to meaning.

Regarding audition, the process starts with the entrance of aural material in the auricle until it transforms into sound waves that reach the brain (Gómez Serrano, 2013, p. 64-70). Nonetheless, perception is also selective, in the sense that we tend to omit what we consider not to be pertinent (Gil Fernández, 1990, p. 69, cited in Bernabeu, 2019, p. 49).

Concerning perception, we must consider that it takes place after audition and it depends on how we notice it. There are many factors that show up in this phase and that may condition how we 'hear', such as noise resistance or audio-visual aids. This last element becomes of great importance for our study, since perception turns into multimodal in the sense that there are stimuli coming from two different means. Thanks to the well-known McGurk effect (McGurk & MacDonald, 1976), it is demonstrated that visual input plays an important role when perceiving sounds, because it can affect the real auditory input emitted. In this case, visual input means looking at the speaker's lips, but it already anticipates that it is not the same listening to only voice than listening while watching. Also related to perception, the following features are to be considered, since, different to written speech, they take part of oral text: background noise, filtered voice (the phone, microphone), unplanned communication, inconstant speech speed, clearness in articulation, non-standard varieties of language (De Santiago Guervós & Fernández González, 2017, p. 492).

d) Difficulties of listening

Listening presents a series of difficulties because of its nature of complexity, oral –and also visual, channels and because many things take place simultaneously and spontaneously. According to Brown (2011), these difficulties arise from a set of factors that listening conveys:

Table 1. Listening difficulties (own creation, adapted from Brown, 2011)

Factors	Implied elements that make the difficulty
Text type	Speech rate, pauses, pronunciation, organisation of the text, live speech
The task	Complexity, level of response required (answer type), level of support (pictures), response time
The speaker	Style, accent number of speakers, recorded speech
The listener	Proficiency level, interest and motivation, confidence, anxiety, familiarity with the topic

2.2.2   Listening with multimodality

Listening not only includes aural inputs but visual. As Jewitt (2013) states, oral communication is multimodal, which means that speech is only one source among other oral and visual sources we use when we build meanings, such as 'sound, music, background noise, gestures, facial expression, body posture, gaze, text, images, figures' (Jewitt, 2013, cited in Campoy-Cubillo & Querol-Julián, 2015, p. 195). Under the command of training learners to be communicatively competent in real life situations, we should, thus, consider multimodal components as a necessary skill to enhance.

Thinking about real life contexts, it should be necessary to look at the contexts that are multimodal, or, instead, it would be more adequate to look at those which are not multimodal, since there are fewer oral communitive contexts which are not. According to Pérez Bernabeu (2019), the only contexts in which sight does not work as a support 'are the radio, the telephone, voice recordings, conversations listened by change and the messages in airports and public transport stations' (p. 57). Also, if looking exclusively at the transactional purpose of possible contexts, those are 'the radio, voicemails, podcasts and audiobooks and recited poems' (p. 58).

Visual elements can be of a wide range of types, and it is not the same if we are referring to speech and seeing the person speaking –which may include articulation, facial expression, gestures–, or speech and audio-visuals, that is, still or moving image, and sounds. This may depend on the nature of the listening material, whether it is a video of a conversation or a more descriptive video explaining a topic with visual aids, that is, looking at people speak or, instead, listening with visual aids. Of course, it can also be a combination of both types. In the present paper, when referring to multimodality from now on, it means audio, still and moving image and written text as modes that contribute to comprehension.

Mayer (2005) suggests the Cognitive Theory of Multimedia Learning (CTML) learning, basing on the dual-channel assumption from Clark & Paivio (1991) and proposes that 'human

information-processing system contains an auditory/verbal channel and a visual/pictorial channel' (Mayer, 2005, p. 33), so this concludes that we do not only listen by a channel but that pictures help in the processing. In other words, Paivio (2007) also states: 'A combination of imagery and verbal information could make L2 input more comprehensible and easily retrievable from memory as the activation of both verbal and nonverbal systems results in better learning'.

Among the numerous studies proving the benefits of multimodal listening, Norte Fernández-Pacheco (2016) highlights the benefits of vodcasts[1]: 'facilitate the introduction or different communicative modes, adjust to students' learning styles, difficulties and disabilities […] and distribute authentic-content situations' (p. 53).

2.3 Listening tasks in the EFL classroom

It was not until the 1990s when there started to be interest in the instruction and evaluation of FL listening (Rost, 2002, cited in Norte Fernández-Pacheco, 2016). Actually, this is one of the main reasons why it was called 'the Cinderella skill' (Mendelsohn, 1994). This name comes from the fact that listening used to be the 'forgotten one' among the 'elder sisters'. Nor should be forgotten that, even in more proficient levels, listening continues to be the skill in which learners tend to be more unsatisfied: 'After passing a long course of instruction, still they have problems in listening comprehension, and may say "I could not understand the spoken texts or any other types of audio materials"' (Serani et al., 2014, p. 31). Even if Rubin's definition (1995, p. 151) includes processing 'visual and auditory clues', it is a fact that no attention has been given to non-verbal sources, at least until recent times, when 'effects of auditory, audio-visual and multimedia inputs on students' comprehension' have been used as new research directions (Norte Fernández-Pacheco, 2016, p. 56). When it comes to multimodality and working in real-world contexts, Serani et al. (2014) claim that 'the presence of audio, in addition to visual information, may convey some sort of feeling to listeners in which they are experiencing a real context with all its details' (p. 31).

According to Nunan (2002), there are different types of listening activities, which foster either top-down or bottom-up processing skills and that help to foster strategies. These are the following: listening for specific information, listening for details, listening for the main ideas, listening for opinions, listening to infer meaning, predicting, and recognising context.

2.4 Previous investigations in the field

2.4.1. Studies on performance with multimodality

---

[1] A vodcast is a video podcast.

There are numerous studies putting forward the use of visual semiotic sources to enhance listening comprehension in EFL learning, and most of them conclude with the positive influence of visual inputs in the overall comprehension of the text (Sueyoshi & Hardison, 2005; Ramírez & Alonso, 2007; Guichon & McLornan, 2008; Wagner, 2007 & 2010; Pardo-Ballester, 2016; Winiharti & Herlina, 2017; Soulaiman et al., 2017, among many others). Some of these studies deserve now to be described in more depth.

Sueyoshi & Hardison (2005) focus on the role of speakers' faces and gestures and, having three types of formats (audio-visual with gestures and face, audio-visual without gestures, and audio-only), demonstrate that all the groups (higher and lower level) perform better in the first two formats. For his part, Wagner (2007 & 2010) compares two groups of students, one with audio-only input and another one with the same input but in a video format. It is concluded that non-verbal information appearing in the video influences students' performances. In the same line of this study, Winiharti & Herlina (2017) come to the conclusion that, having two groups with the same listening material, one with the screen on and the other with the screen off, it is the group that has access to video clues that obtains better scores in a comprehension test based on a specific number of items:

> Through the videos, they can perceive other kinds of information –pictures, moves, gestures and writings. The findings then suggest that the exposure [sic] to more than one modality will bring students to better performance in listening comprehension (p. 191).

Soulaiman et al. (2017) also reach the same conclusion, and emphasise the reason for this to the fact that video provides with the use of meaningful and real-life situation contexts and language:

> Based on the results attained, it was found that the majority of the students obtained higher marks when using video media method compared to audio only method. Hence, it is recognised that by using video as one of the assessment tools will help students to perform better due to the use of authentic, meaningful and real-life situation contexts and language (p. 201).

Guichon & McLornan (2008) also agree that comprehension improves –performance is higher– when learners are exposed to text in several modalities (focusing also on subtitling incorporation), while making participants write a summary of the listening comprehension material under the context of an experiment. Nevertheless, these authors also claim that, even if image incorporation normally reinforces comprehension, sometimes, if the visual information is not directly connected to what the aural information is saying, it can entail a cognitive load and distract attention when processing (p. 8). Pardo-Ballester (2016) also has some objection with some drawbacks that multimodality can involve, especially in the lower stages of FL learning. In her study, where redundancy is used as a means of emphasizing on the elements that want to be tested in unimodal and multimodal format, she concludes that beginners seem

to perform worse when inferring meaning in listening with video format, and she attributes this phenomenon to the possible fact that learners have not been very exposed to language and because they are not used to listening, watching and paying attention to written elements to be answered. This evinces that listening is, thus, a more complex process when visual content is present and should be kept as a methodology at higher levels.

All the previous studies support that learners score better when having visual aids, although some other investigations have not come up with conclusive results that can make such claim (Gruba, 1993; Coniam, 2001, & Batty, 2015). For instance, Batty (2015) finds too insignificant differences with only a slight facilitative effect of visual cues. The author remarks that the test used was 'according to a test specification that stipulates that items must be answerable from the audio content alone', so 'the test likely does not take advantage of the extra information available from the non-verbal channel' (p. 15).

2.4.2. Studies on attitudes towards multimodality

Not only is it important to analyse the performance, that is, the results that learners obtain in oral comprehension tasks, but to ask the learners themselves how they feel about the incorporation of multimodal content when doing listening comprehension. Some studies combine both performance and attitudes regarding multimodal channels. An example of this is Pardo-Ballester (2016) and one of the main conclusions she comes to is that, regardless of the performance of each of the learners in the tests, video is always their choice of preference. Norte Fernández-Pacheco (2016) focuses her investigation on the students' attitudes towards the application of vodcasts (a multimodal format) in the EFL classroom, comparing it to the traditional audio-tracks. Under the parameters of 'sound quality, length, speed, enjoyment, difficulty, anxiety and language improvements', the results through a questionnaire indicate that the attitudes are always positives and help students improve comprehension as well as enjoyment and motivation for learning the language.

The outlook to the already existing literature on the field of listening and multimodality helps in the design of my own investigation and anticipates some of the outcomes that I might come to. I position my investigation similar to the others taken place in EFL teaching, but my focus is on Post-Obligatory Secondary education (Batxillerat).

2.5 Research questions

As stated in the objectives, this study pursues to obtain two types of response: on the one hand, the aim is to verify whether it is true that oral comprehension[2] is higher when the input consists of listening and watching (LW), rather than when it consists of listening only (LO). The focus of the analysis is performance (students' results). On the other hand, a second aim of this study

---

[2] 'Comprehension' is somehow a biased concept by itself, since comprehension performance will be based on specific questions that may leave some information aside.

10

is to analyse the students' attitudes towards multimodal materials, and, as far as possible, seek the reason for a preference. Thus, the focus is their perception.

Once having presented all the necessary concepts and previous research in this field, these are the research questions to which the study aims to respond:

1. How is learners' comprehension different in terms of performance between LO and LW tasks?

2. Can listening and watching materials (LW) have an effect on improving learners' comprehension?

3. What are learners' attitudes and perception towards the application of multimodal listening tasks (LW)?

3   METHODOLOGY

This study combines qualitative and quantitative research methodologies. On the one hand, it aims to examine quantitatively if there exists some difference in comprehension in LO and LW tasks and if LW materials can increase comprehension. On the other hand, the methodology also aims to examine both quantitatively (degree of agreement) and qualitatively (open-ended questions) students' attitudes and perception towards LW.

3.1   Participants

The sample for this investigation consists of twenty-two 1st of Batxillerat students who belong to Institut Rafael Campalans (Anglès, Girona). They are all between 16 and 17 years old, with 14 women and 8 men. The group is a mixed ability class and not only includes students whose L1 is Catalan, but students with origins from all over the globe: two students from South America, two students from Morocco and one student from India. The reason for the choice of students in this level is derived from two premises: (1) students Batxillerat are more mature, and (2) because of their level of English. As Pardo-Ballester (2016) indicates, debutant learners may find it too complex to listen, watch and answer items at the same time because of little exposition still to the language. The level that the subjects in this study are following during the present intervention is B2.2. of the CEFR, although only a few of them have obtained an official certificate in this level.

A piloting group has been used in order to guarantee that data-collecting tools are understandable and clear. This group is the other 1st of Batxillerat group (same level). Results from the pilot study have helped in the design of the tools, especially to ensure reliability between the two listening comprehension tasks.

3.1 Tools

The study used three different instruments to retrieve data: two performance tasks to measure students' performance after carrying out two listening tasks and a questionnaire to retrieve students' perception on the task. While tools 1 and 2 are exclusively quantitative, tool 3 shows a mixed methodology. It uses numbers, e.g. in the degree of agreement, but part of the data can also be analysed qualitatively (open-ended questions, where answers emerge from the students). Quantitative data is analysed with descriptive statistics whereas qualitative data is analysed with content analysis and ranking according to frequency.

3.1.1 Tool 1 & 2: Listening tasks (LO and LW)

In order to gather data on the performance of oral comprehension tasks and compare the results in LO and LW, two different listening tasks had to be designed: a podcast, which was used for LO and a Youtube video, which was used for LW. They share the topic of soda taxes, but the content of each piece is slightly different. They both are natural and authentic materials, giving real-life contexts to learners. The length of both of them is 3 minutes. Their speech rate has been calculated and it stays on the average rate (140 wpm[3]), which is neither slow nor fast, in both cases. The accent of both speeches is American English. In both cases, the purpose of the listening is transactional, which means that there is no interaction and there is a single speaker (Vandergrift, 2011).

1. **LO, taken from 3-minute podcast on PlayerFM and accessible also on this link: https://www.youtube.com/watch?v=ODS2t6hCV2I**

This audio track is a short podcast with only a speaker, who we infer to be a woman, who weekly talks about some issues. In this case, under the title 'Do soda taxes work?', she offers a monologue describing taxes that have been applied to sugary drinks and criticises why these taxes do not have the same effects as taxes in tobacco had.

2. **LW, taken from WFPC channel on Youtube and accessible also on this link: https://www.youtube.com/watch?v=3fNYw0CVfl4**

This consists of an informational video by Duke Sanford (World Food Policy Center) talking about the sense it makes for our health to tax sugary drinks. There is a speaker, who we infer to be a man. It is multimodal in the sense that there are audio-visual elements all along the format, but no real people show up, which means that semiotic sources such as facial expressions and gestures are not present in this case. In contrast, there is background music, moving and still pictures and icons, symbols (such as 'ticks') and many written words. Element transactions are fast, as the speaker talks about the different issues.

---

[3] These data has been calculated on Debatrix.com

Each task included a comprehension questionnaire which were tailor-made from scratch for this study. The question types are either multiple-choice or short-answer, and possible bias in random answering concerning multiple-choice are fixed by incorporating distractors.

Both listening comprehension tasks consist of 14 items, making them easier to compare in the quantitative analysis. It was ensured that the type of response between the two listening tasks was balanced: LO comprehension includes 6 multiple-choice and 8 short answer questions, while LW comprehension includes 5 multiple-choice and 9 short-answer questions. The two listening tasks that participants had to answer are available in Appendix 1a (distractors indicated) and the answers can be found in Appendix 1b.

It is worth to bear in mind that all the questions can be answered from the aural content of the material, so visual elements in the LW task work as support and most of the time. After the piloting, it was decided that the tasks were going to be played three times as a fair number for students to answer successfully.

Before the tasks, a pre-task was carried out, framing this 'experiment' in the structure that an ordinary listening task in the classroom context would have. The topic between the two tasks was shared, so a collective pre-task activating previous knowledge on 'sodas', 'sugar' and taxes in the first session was enough for both listening tasks (See pre-task in Appendix 1c).

3.1.2 Tool 3: Questionnaire about the students' attitudes and perception towards LO and LW

With the aim of investigating students' attitudes and perception towards multimodal materials, a questionnaire with quantitative and qualitative questions was used.

The questionnaire starts with background information (section 1), such as name, age, gender, and if the student learns English somewhere else apart from the lessons there. In section 2, participants are asked how frequently they practise different listening habits: three items refer to LW format (films, Youtube and videogames) and other three, to LO (music, podcasts and the radio). The aim of these questions was to know what they do in their free time. It was not asked whether they were more exposed to LO or LW listening tasks in class, since this was something that the teacher already knew. The book they are using this year does not have any multimodal listening track, since it follows the Cambridge examinations' (B2.2) structure, and, therefore, multimodality in listening rarely appears.

In section 3 participants state their level of agreement to a set of statements. Using a Likert scale, from 1 ('strongly disagree') to 5 ('strongly agree'), different statements cover 9 parameters: (1) sound quality, (2) length, (3) length, (4) speed, (5) entertaining, (6) difficulty to understand general information, (7) clarity of articulation, (8) anxiety and (9) recalling information. Statement always have the same structure ('LW –x parameter- than listening LO').

However, in order to avoid transmitting biased opinions with the statements, a combination of positive and negative senses of the characteristics are used, so that statements are not done in favour of one of the two modalities. In the case of items 6 and 7, in the analysis, values have been reversed in order to make it easier to interpret together with the other data (e.g. the statements 'it was more difficult to understand…' and 'it is less clear…' have been turned into 'it is easier to understand…' and 'it is clearer…').

It is worth mentioning that items 1, 2, 4, 5, 6 and 8, in section 3, have been taken from Norte Fernández-Pacheco (2016), whose study also analyses attitudes towards multimodal listening. The other items have been added because they were also relevant to analyse in this study. Item 7, which is related to item 6 on 'general comprehension', has a nuance of 'understanding through intelligibility', that is, word by word. Item 6 has similarities with top-down or bottom-up processing, while item 7 only has to do with bottom-up (more detailed listening). The other parameter (item 9) is the ability of 'retrieval' (recalling information) and, since tool 1 cannot check vocabulary acquisition, this item is a way to ask if the participant feels that that format allows him or her to remember information more easily.

Section 4, with open-ended questions, asks for preference and language improvement and the reason for that (See Appendix 2).

3.2   Procedure

The experiment took place in the ordinary class of the participants in Institut Rafael Campalans (Anglès), with 22 students in test conditions, sitting separately. Each student had his or her laptop and used earphones. The listening input was on their screens and the listening comprehension questions were administered on paper support. Participants' data is anonymous, although a number is given to each of them ('subject *1-22*') in order to see if there is a relationship between performance in tool 1 and attitudes and perception in tool 2.

The two listening comprehension tasks took place in two consecutive days. On Monday 22nd March 2021, a pre-task was carried out (see section 3.1.1), and 11 students did the LO (Type A) and the other 11 did the LW (Type B). On Tuesday 23rd March 2021, the types were reversed, what made tool 1 and 2 conclude after the two tasks carried out. The same day – second day–, tool 3 was carried out and all the students filled in the questionnaire on Google Forms.

3.2.1   Groups: counter-balanced methodology

A counter-balanced methodology in the design of the experiment was needed for tool 1 and 2, so the students were randomly split into two groups (eleven students each). Group 1 (subjects from 1 to 11) is the group doing first LO and then LW, while group 2 (subjects from 12 to 22) is the group doing first LW and then LO. This way, we ensure that many factors that could affect the results were reduced: we might not be able to control factors of the context of the student him/herself or the time in which it takes place, but, at least, we avoid that having done

14

one of the formats first may have some impact in the results. With a random division of the groups, it cannot be controlled that in some group there are more proficient students than in the other. This counter-balance lets us compare the difference between one mode or another (LO and LW) and the difference between the two groups of the sample.

3.3 Analysis

Regarding tool 1 and 2, descriptive statistics were used for analysis. Average values and percentages are important to conclude which format fosters a higher level of comprehension and standard deviation indicates how responses differ from the mean. Out of a total of 14 items in each listening comprehension task, each answer was given a value of 1 if it was 'right' and 0 if it was 'wrong'. Concerning tool 3, descriptive statistics were also used to analyse whereas content analysis was used.

4 RESULTS

4.1 Results in the performance (tool 1 and 2)

Counter-balanced methodology

Group 1 scored 50.6 % in LO (3.8) and 73.3 % in LW (1.8), while group 2 scored 62.3 % in LO (2.5) and 84.4 % in LW (1.3). Even though the standard deviation indicates that there are individual differences in each group, the difference in performance among groups (1 and 2) is not significant. In Figure 3, it can be seen that group 2 has higher scores in comprehension for both LO and LW. In particular, the percentage of comprehension of group 1 is of 22.72 % higher in LW, while it is of 22 % in the case of group 2, so the difference in comprehension does not even reach 1 %. This leads to the conclusion that the order the tasks are taken does not influence performance.

Table 2. Mean scores per group

	Scores LO		Scores LW		Difference between formats (in favour of LW)	
	%	Mean score (SD)	%	Mean score (SD)	%	Mean score
Group 1 (LO + LW)	50.6 %	7.0 (3.8)	73.3 %	10.2 (1.8)	22.72 %	3.1
Group 2 (LW + LO)	62.3 %	8.7 (2.5)	84.4 %	11.8 (1.3)	22.0 %	3.0

Performance (groups and overall)
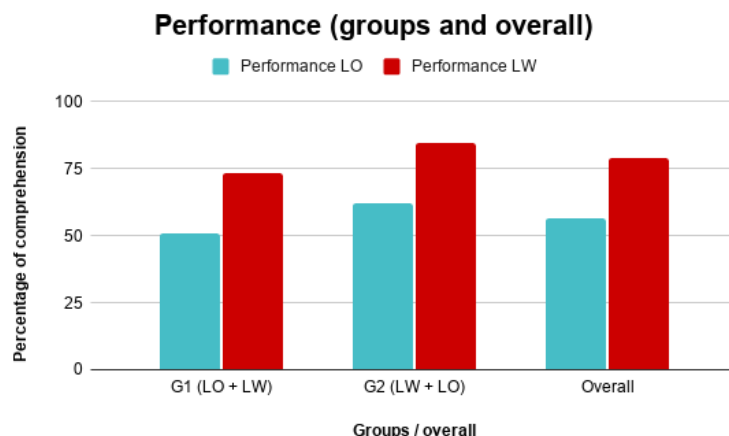
Figure 3. Percentage of comprehension in LO and LW tasks

Comparison performance by subject
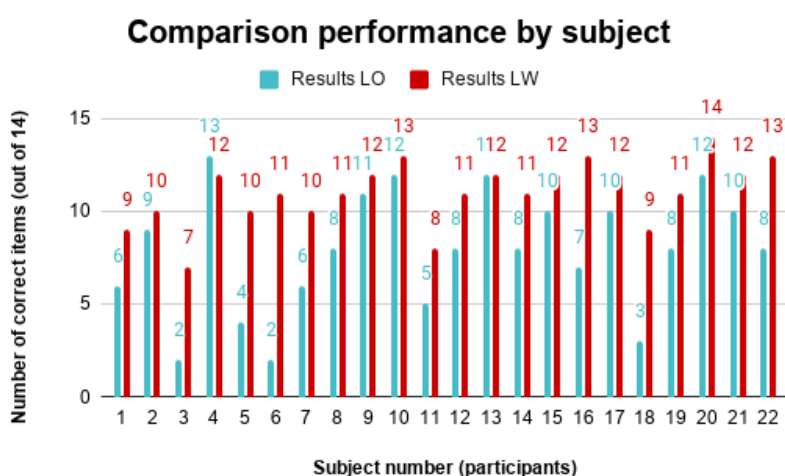
Figure 4. Performance by each subject in each task

LO vs. LW performance (overall)

Once having showed that the order is not relevant, LO and LW can be considered together, regardless of the order that the listening tasks were performed. As Figure 3 above and Table 3 below show, the overall performance percentage is of 56.4 % for LO task and 78.8 % for LW task, which means that comprehension increases 22.4 % more on average when the listening is in a multimodal format. As a general datum, 90.9 % of the participants have obtained a higher score in the listening and watching format (LW) and, thus, it can be claimed that comprehension increases in this task when there is a visual input in the listening material.

Table 3. Mean scores (overall)

Scores LO		Scores LW		Difference between formats (in favour of LW)	
%	Mean score (SD)	%	Mean score (SD)	%	Mean score
56.4 %	7.9 (3.2)	78.8 %	11.0 (1.7)	22.4	3.1

16

Regarding the items in the task separately, out of the 14 items in the tasks, on average students score 3,1 items higher in LW. If we look at Figure 4, it is relevant information that some students improve their performance enormously when listening and watching to the video (LW). Even if some subjects only have a difference between LO and LW of 1 item (7.14 % more in favour of LW), the difference between the two modes can reach 9 items (64.28 % more in favour of LW). This is the case of subject 6, who passes from 2 items guessed in LO task to 11 items in the LW task. Among the total of students performing higher in LW, 65 % obtain a higher score comprising from 1 to 3 values higher (7.14 -21.42 %), while 35 % do so but with scores comprising from 4 to 9 values higher (28.57-64.28 %). Conversely, only 9.1 % of the participants show lower scores in LW. That is, one participant (subject 4) guesses more items in the LO format, while one participant (subject 13) stays the same in the two formats (see Figure 4). Standard deviation values in Table 2 show that there are individual differences in performance. While in LO format standard deviation is of 3.2, in LW it is only of 1.7. This means that scores variate more in LO formats, while they stay more constant among subjects in the case of LW.

Hence, it can be summed up that all the students' scores, with the exception of 2, are higher in the case of the multimodal material tasks, with little or very significant difference gap (see detailed scores per participant in Appendix 3a).

Analysis of each item of the tool

Regarding the results that can depend on the design of the task, it should also be considered that there is also a great variability between those items that everyone guesses right and those that only a few participants score right. This means that there are certain items in which there is a tendency to either collectively guess right or, instead, fail. Figures 5 & 6, where we find performance per item, whose values comprise from 0 to 1, with 0 standing for 'wrong' and 1 standing for 'right' (see detailed item by participant in Appendix 3b) show those items which are commonly guessed, or those in which there is a common tendency of guessing it wrong. It is relevant to see that, in the case of LO, this variability is wider, depending on the question, while in the case of LW, value of guess is much higher most of the time, with three little exceptions: items 5, 6 and 14.
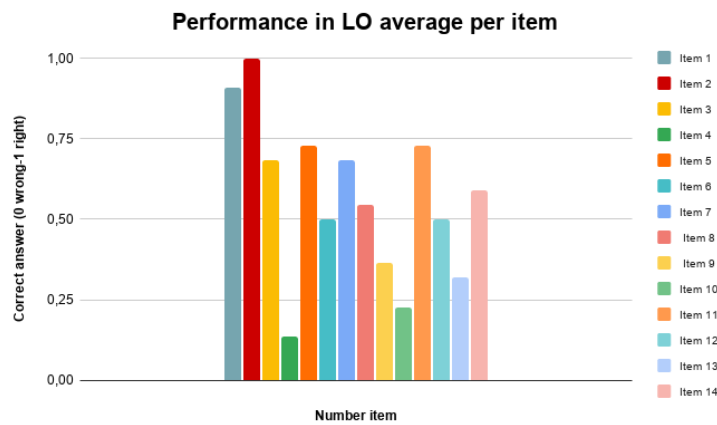
Figure 5. Overall performance per item (LO task)
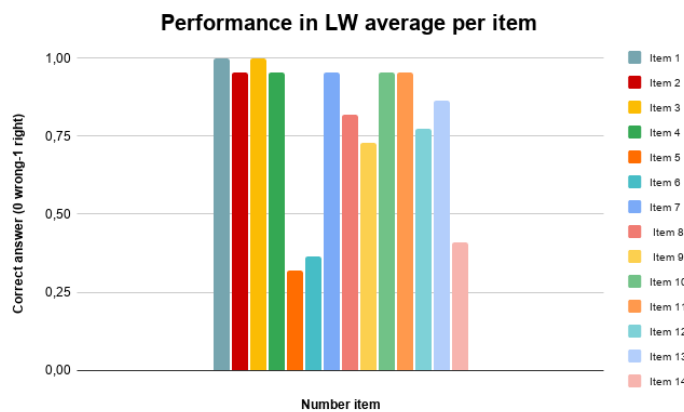
Figure 6. Overall performance per item (LW task)

4.2 Results in the attitudes and perceptions questionnaire (tool 3)

In this section, we look at the attitudes that emerge from participants regarding the multimodal and unimodal formats based on their experience in LW and LO tasks respectively. Most of the data is analysed quantitatively, except for the two final questions which cover students' opinions on mode preference and language improvement.
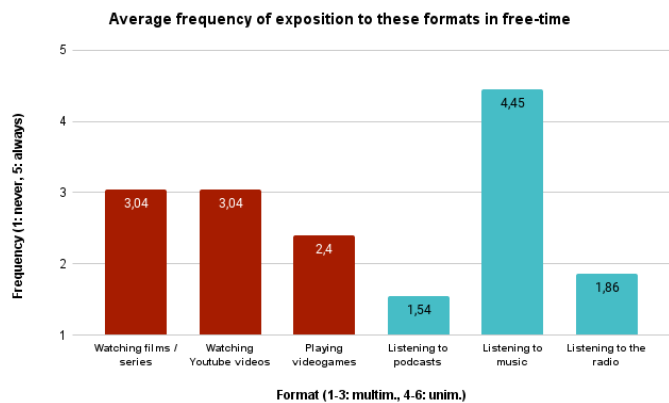
Figure 7. Average frequency of exposition to multimodal (1-3) and unimodal (4-6) formats

This analysis starts with a view to which formats students claim to already be exposed in their free-time to see if there is a tendency of students using more unimodal or multimodal sources to practice English outside the classroom. It is found that the most spread and constant format for learning English is 'listening to music', with mean frequency of 4.45 value of frequency, extremely close to 5 ('always'), and with a value of 0.67 in standard deviation, being the littlest value distribution in the set (see Figure 7 –with LW in red and LO in blue– & Appendix 4a).

18

Music is a LO format and it is the most used one, although it is counteracted by the values they give to 'podcasts' and 'the radio', which are the lowest in the overall frequency. Conversely, if we look at the first three columns, corresponding to LW format, the averages for 'watching films' and 'watching Youtube videos' stand just over value 3 ('sometimes'), with a standard deviation of approximately 1 in both cases (1.29 and 1.09 respectively), meaning that there is a variety of exposition in the group. In the case of videogames, the average stands lower and it is the format in which standard deviation value is higher (1.46), meaning that only subjects 2, 9 and 10 claim to 'always' play videogames, while many others state to never play them, with value 1. Finally, it was interesting to know if students are exposed to more English outside the classroom, 9 participants claim to learn English at an academy, 3 claim to have attended English lessons in the past but not now and 10 claim to have never attended English lessons outside the classroom.

The degree of agreement with the statements comparing parameters of each format is the following (see specific values answered in Appendix 4b and 4c):

- Regarding sound quality (it. 1), 90.9 % of the students either strongly agree or agree that LW has a better quality of sound.

- When it comes to length, comparing the two items (it. 2 and 3), it can be stated that LO is the format that students find longer, with 45.4 % agreeing or strongly. It is a feature in which more than a third of the students stay neutral and, regardless of variety in the rest, no one claims to find LW long.

- Concerning the speech speed (it. 4), 59 % declare that speakers in LW format spoke slower, while 13.6 % have the opposite feeling.

- In reference to entertainment (it. 5), 90.9 % see LW as more entertaining than LO format, with an average of 4.45 value of agreement, being, together with item 9, the most agreed parameter in favour of LW.

- Regarding facility to understand general information (it. 6), 90.9 % agree or strongly agree that it is easier to understand general data with multimodal aids (LW). 9 % stay neutral, and no one (0 %) disagrees or strongly disagrees.

- When it comes to clarity of articulation (it. 7), 86.36 % agree or strongly agree that they can understand more clearly what is said word by word in LW.

- Concerning anxiety (it. 8), attitudes are diverse, although the average response is a value of 3.77 and 72.7 % of the participants concentrate in the block of agreeing or strongly agreeing.

- In reference to retrieval (it. 9), it is the one of the more agreed characteristics, with 95.4 % of the sample stating that in LW it is easier to remember information than in LO.

To sum up, some of the most agreed parameters are that LW has a better sound quality, LW is more entertaining, in LW it is easier to understand general information, and easier to remember

information. Value distribution among the participants is smaller in 'easier to remember information' and 'retrieval', showing, therefore, that there is great consensus in attributing these characteristics to multimodal listening format. Length of LW (it. 3) and anxiety (it. 8) seem to be those where attitudes are more diverse.
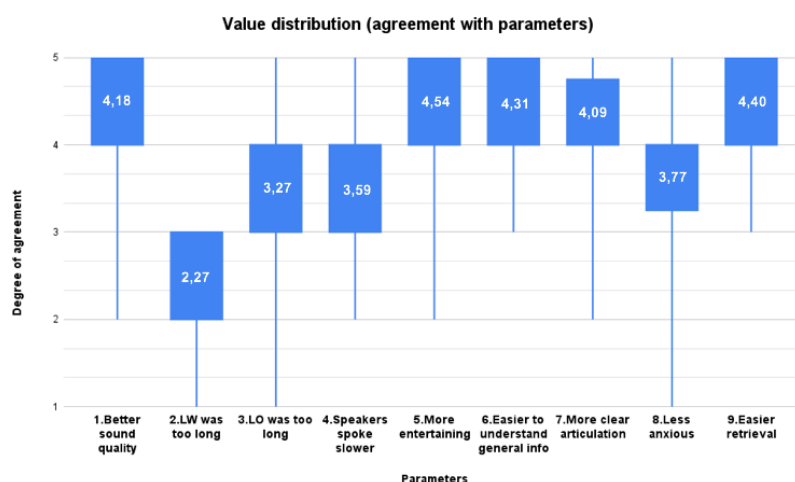
Figure 8. Boxplot with value distribution and mean (1-5)

In the open questions where students had to explain their mode preference, 100% of the participants stated that LW is the format they prefer for carrying out listening tasks. Nevertheless, concerning possibility of language improvement, there is not total consensus and 81,82 % believe that LW will help improve their level of language. Now, two tables collecting students' reasons for each of the two questions will be given. Most of them are repeated among the sample, so they are listed in order of frequency of presence in students' responses.

Table 4. Reasons for supporting preference for LW (ranked by percentage)

- There are many images and some words were written. This helps to know the answer. / The main ideas were drawn and you could read words **(26.66 %)**.
- It is more entertaining / appealing, and I enjoy it more **(16.66 %)**.
- Even if I could not understand every single word, I could understand what the speaker was talking about. / It is easier to deduce meanings **(16.66 %)**.
- There is more information than in an audio track and I can understand specific words, such as the name of diseases at the beginning of the video **(10 %)**.
- You can retain information. / You can recall images **(10 %)**.
- The speakers spoke more slowly **(6.66 %)**.
- I could follow more easily / I could understand the structure of the information **(6.66 %)**.
- You keep more focused / concentrated **(3.33 %)**.
- It is more similar to the videos I sometimes watch **(3.33 %)**.

In Table 4, we can see that the most recurrent reason is that visual aids –image and written text –in the video help in comprehension in tasks. Also, students highlight among preference, 'more enjoyable content' with multimodality and the fact that 'meaning deduction can be made'. In

20

less degree, there provide other reasons for their preference, such as 'specific words access', 'information recall through image', 'slower speech', etc.

Table 5. Reasons for language improving with LW (multimodality)

FOR	AGAINST
• I can suppose information through images, so that I do not get lost. / I can make association of words that I do not know **(30.76 %).** • I have the feeling that the speaker speaks more slowly **(15.38 %)**. • It is easier to remember thanks to the pictures **(11.53 %)**. • It motivates me more to listen and to continue learning more **(7.69 %)**. • Context is understood more easily **(7.69 %)**. • I feel more focused/concentrated **(7.69 %)**. • I do not frustrate as much as with audios **(3.84 %)**. • Audios have a part of interpretation that in the video is reduced (I do not need as much knowledge in the topic) **(3.84 %).** • I do not need to concentrate as much as with audios **(3.84 %)**. • It is more entertaining **(3.84 %)**. • It is closer to reality **(3.84 %)**.	• In an ordinary conversation in English you will not have a visual support (the person will not give you a video), so it is a good idea to get used to relying to your listening skills without it **(50 %)**. • Even if it can be more difficult to understand, audio will help me, since, if I understand audio, it will be easier to understand other types with images **(50 %)**.

In Table 5, we can see a very recurrent reason for multimodality to improve language, which is 'word association to unknown words' and 'image as support in comprehension'. 'Slower speech' is also a feeling and 'information retrieval', again, like in Table 3, too. Also, students assert 'motivation', 'context understanding' and 'more concentration'. A single individual reports 'less frustration', 'less interpretation needed', etc. As for the reasons against multimodality, there is consensus among students claiming LO to be more useful for improving language in that 'not always will they have the opportunity to have visual support when listening' and that 'being more trained in LO listening is better for later being able to understand all formats'.

4.3 Discussion

Once having analysed the collected data, it is now time to discuss what these data imply, and drive us to draw a firm conclusion of the objectives that this research pursuits.

Learners' comprehension in LO and LW tasks

Regarding participants LO and LW comprehension, it has been claimed that comprehension in the present study increases in 90 % of the cases when the listening input consists of listening and watching at the same time, with higher or lower variability between modes depending on the subject. The fact that 65 % of the participants showed only little variability between the two formats (from 7.14 % to 21.42 % higher in LW) indicates that multimodality does not always involve greater comprehension, but the fact that the other 35 % show extremely different

performance (from 28.57 to 64.28 % higher in LW) mean that, for them, multimodality can be a good way to boost their comprehension, as studies seen in literature have stated too (Sueyoshi & Hardison, 2005; Winiharti & Herlina, 2017; among others). Also, a counter-balanced methodology was justified to be applied, since data demonstrates that, regardless the order in which listening tasks are carried out, performance does not change at group level.

In terms of the tool design, it is important to take into account that, different to what some studies have done, in this research it was decided not to alter the natural format of the listening file, as means of preserving authenticity of the formats. For this reason, this study does not follow the same process as Wagner (2007, 2010) and Winiharti & Herlina (2017) have done applying the same input in two different modes. In this study, it was thought that altering a video to audio-only format might have had some effect and would not keep natural formats of the chosen materials.

Even if performance of the listening task is generally higher in LW format, it seemed relevant to examine the most guessed and the most failed items in the 14-item tasks. In general, it can be established that very specific questions containing the explicit words both in the input (either spoken or spoken and shown with text or image) and in the questions of the task are those in which participants have scored higher, opposed to others which implied a higher cognitive load in information, such as a lot of visual and aural elements at the same time, or questions were not as explicitly connected to the correct answer.

Looking back at Figures 5 & 6 and tables in Appendix 3b, it can be seen that in LO task, items 2, 1, 5, 11, 7 and 3 are those in which guess value is higher, between 1 and 0.68. Items 1 and 2's answer was 'sports' (1) and 'energy drinks' (2), which was explicitly said in the audio track. Item 5's answer was 'obesity', which can lead us to think that words which are more transparent from the L1 and are very closely related to the topic of the listening too are more understood and, thus, comprehension is more guaranteed. It happens the same with item 3, whose answer is '(type 2) diabetes'. Instead, the most failed items are items 4, 10 and 13, with values from 0.13 to 0.31. In the case of item 4, which is concentrated in the same question of items 3 and 5 just mentioned, in this case the answer is 'tooth decay', which can make part of the topic but the piece 'decay' was not guessed right because it is not as transparent as the others were. Item 10 also seems an item in which participants struggled since 'in restaurants' was a very concrete fragment hidden between a long speech talking about several other things and, even if 'in what place' was explicitly asked as a question, did not help most of the students identify it as the answer. Item 13, for its part, also implied subjects to reverse what was explicitly said to what was said in the other way around in the question in the task.

The effect of LW

Regarding LW task, different to LO, a good number of items are widely guessed right (from a value of 0.72 above), while only 3 specific items show very low performance. The most guessed items are 1, 3, 2, 4, 7, 10 and 11, with a value between 1 and 0.95, which means that all the

22

participants guessed it right (items 1 and 3) and the rest everyone guessed it right with an exception of a single individual who answered wrong. Items 1, 2 and 3 consisted of literally coping the name of the diseases that were shown written on the screen, so students grasped the words. Something similar happened with items 10 and 11, 'alcohol' and 'tobacco', in which students had transparent enough words and they were also showed in visual input (still image). Instead, the most failed items are 5, 6 and 14, with values from 0.31 to 0.40. Even if items 5 and 6 are answered from the perspective of having a lot of visual information explicitly shown or written in the video, what has done students to generally fail these items might be the fact that in that part of the video there is a lot of text and speaking, saying it in other words, at the same time, creating, thus, a cognitive load that some students might have found more difficult to comprehend than if they had had only one of the semiotic sources itself. Actually, as Guichon & McLornan (2008) assert, 'if visual information is not directly connected to what the aural information is saying, it can entail a cognitive load and distract attention when processing' (p. 8).

From the point of view of the counter-balanced design, which has demonstrated that the order of the task does not affect performance, it is to say that in item 3 of the LO task ('obesity'), it can be seen that in this case performance was indeed affected, since group 1, having conducted LO first, scores average value 0.45 for this item while group 2, doing it as the second task, does so in a 0.90 value. This can be explained due to the fact that 'obesity' was also a word already explicitly shown and asked in the task of LW (item 1). Although it was dismissed to assess vocabulary acquisition for this research, this issue can be some evidence that multimodal listening can provide listeners with input that, the second encounter with that word (in this case, later, in unimodal listening), students will recall the word.

Some previous investigations (Gruba, 1993; Coniam, 2001, & Batty, 2015) have not been able to put forward conclusive results regarding the increase in comprehension when having also visual aids. Even if, in the case of our study, conclusions seem to be more similar to those studies supporting its benefits, what is true is that what Batty (2015) uses it as an argument to discard multimodality the content of the tasks was answerable from the aural content only, so we have not been able to prove how different modes play a role by themselves, but only as a combination that can entail redundancy of what is articulated in voice.

Students' attitudes and perception

Concerning research question 3 about attitudes and perception of the listening tasks of the two formats, data is also in favour of the multimodal format, with more positive attitudes broadly extended among most of the sample. All students show positive attitudes in favour of multimodal listening, establishing that LW listening –at least in this specific task– has a better sound, seems shorter to listen, its speakers seem to speak slower, is more entertaining, it is

easier to understand general information, articulation is understood more clearly, causes less anxiety and it is easier to remember information. No matter the performance, there is consensus in attributing these characteristics to the multimodal mode and show preference for it, as it happened in Pardo-Ballester (2016). The most agreed parameters, with over 4,31 average value, are, in order of strong agreement, the following: entertainment, easier comprehension of general information, easier recall of information, and sound quality. With this, it can, thus, be claimed that participants find a big difference in terms of entertainment between LO and LW. It has to be born in mind that participants belong to the generation of the digital natives, so they have grown up surrounded by technology and the image breakthrough in the digitalisation era, as Prensky (2001) suggests. The same idea can be attributed to length, since, as students claimed in section 2 of the questionnaire, they seem to be more used to watching films/series and watching Youtube videos, opposed to LO formats such as podcasts and the radio. Also, as they then claim in the qualitative answers of the questionnaires very recurrently too, they strongly consider that LW implies easier recall of information and easier understanding of the overall speech, which can be attributed to clues apart from the voice (image and explicit words highlighted). This confirms that data collected in the qualitative parameters matches with their qualitative responses when asked for preference and its reason. Some of the most recurrent answers emerging from students themselves are the benefits of association and supposition that multimodal sources can bring for language improvement, something that seems more difficult for them when only having audio input. They also assert to have the feeling that the speakers in the video talk slower and that it is more motivating to continue listening and learning when having video. Some of them also claim that LW helps them keep more focused in listening and completing the task. Nonetheless, almost 20 % of the participants also affirm that video will not help improve language since they are aware that, in either real-life or other tasks, they will sometimes have to rely only on unimodal formats, which means that, although they prefer video, they are more critical and consider other contexts in which multimodality will not be possible. As Pérez Bernabeu (2019) established, those contexts in which sight does not work as a support 'are the radio, the telephone, voice recordings, conversations listened by change and the messages in airports and public transport stations' (p. 57).

It has been demonstrated that listening and watching tasks enhance comprehension compared to listening only tasks. Not only it concerns performance but multimodality is the preferred mode by all the participants, establishing positive attitudes towards LW tasks.

5 CONCLUSION

5.1 Obtained data

The present study aimed to investigate if learners show any differences in comprehension depending on the number of modes that listening tasks contain. In order to accomplish the first and the second objective of the study, once having presented what multimodality can contribute

in communication, two tasks have been undertaken by participants to measure comprehension when the input consists of only audio or when, in contrast, the input consists of a video with three types of sources of information: audio, image and written text (LW). In the same line of previous research, in the present study it has been demonstrated that students' performance increases more than 22 % on average when learners have visual sources apart from voice only. In order to guarantee that carrying out one of the tasks before the other did not affect overall performance, a counter-balanced design has been used. Comparing the two groups' performance has shown that the order does not affect the results they obtain, that is, their comprehension. Although it can be claimed that learners seem to generally score right those items in which there is explicit connection of what is said (audio) and what is shown (image and text), in a few cases, as some studies pointed out, it can sometimes imply a high cognitive load and result in inconvenience. In short, hence, it has been shown that the use of different modes influences the way students perceive and understand information.

In reference to learners' attitudes and perception, which is the third objective of the study, participants have shown positive attitudes and perception towards multimodal tasks, especially for its features of entertainment, easier comprehension of general information and easier recall of information, to which we can add, as answered qualitatively, easier deduction and supposition of the information. They overwhelmingly support the advantages of multiple modes against a single mode and show unanimous preference for multimodality. More than 80 % of the sample also considers listening while watching as means to improve their language level. The fact that students would show a more positive perception of the multimodal task is a hypothesis that, as proved in other research, has recurred also in this study.

Regarding usefulness of this investigation for language learning and language teaching, once having reached these conclusions, it is important that language teachers consider the numerous benefits that multimodal inputs and tasks can offer to learners in the learning process. With what it has been seen with the most extreme differences in some students' performance, teachers need to bear in mind that listening and watching both enhances comprehension and is closer to students' current learning environments and habits. In other words, as Soulaiman et. al (2017) claimed, the reason of the success of multimodality is that it uses authentic, meaningful and real-life situation contexts and language.

5.2   Limitations of the study and further investigation

As limitations of this study, the most relevant one is that data is not representative enough because the sample is based on only 22 individuals, which seems to be insufficient to extend it to the whole population. However, the present study can draw us lines to research in the future with a more representative sample. Also, the tools –listening tasks– for assessing

comprehension were tailor-made from scratch and, although a piloting group was useful to check that the task was reliable and clear to participants, we ensured students to have three times to listen to the input, in case the task was difficult for students. If compared to the fact that test conditions tasks sometimes give only two chances to listen, our study considered a third listening after having conducted it with the piloting group.

The fact that the methodology has not been designed to test vocabulary acquisition is also a restriction of the research, since multimodality has often been linked to how learners can learn vocabulary, as Hsieh (2019) and Tavakoli & Gerami (2012) do in their studies. Nonetheless, as it has been pointed in the discussion, participants who came across the word 'obesity' for the second time (in LO) scored right with a higher percentage than those who had perceived it in only audio in the first listening, which means that both multimodality and encountering a world for a second time helps acquisition and retrieval.

Something that this study does not include is gestures and face, different to what Sueyoshi & Hardison (2005) do when comparing the formats of audio-visual with gestures, audio-visual without gestures and audio-only. If the video chosen for the task had included human people speaking, instead of moving and drawn still pictures, other semiotic sources, such as gestures and face expression, could have been tested and used as context. The same has happened with the purpose; the audio and video chosen did not have an interactional purpose but a transactional one (Vandergrift, 2011), so the research only focused on one speaker transmitting a message. It would have been interesting too to see the result a two-way task, such as a conversation, would have provided.

The limitations just mentioned above evince considerations which could be made in the future to continue the lines that this investigation opened. This study concludes with higher comprehension when multimodality is applied and, according to participants' perception, it helps them improve language. The research is only based on a single intervention, so a step to take in the future would be tracking if using multimodal listening tasks in the long term shows significant progress and improvement in learners.

6 REFERENCES

Bezemer, J. & Jewitt. (2010). Multimodal analysis: key issues. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 181-197). London: Continuum.

Bezemer, J. & Kress, G. (2014). Young people, Facebook and pedagogy: recognizing contemporary forms of multimodal text making. In M. Kontopodis, C. Varvantakis, M. Dafermos & C. Wulf (Eds.), *Youth, tube, media: Qualitative insignts and international perspectives. Berlin: Waxmann.* Retrieved from https://www.academia.edu/6916873/Young_people_Facebook_and_Pedagogy_Recognizing_Contemporary_Forms_of_Multimodal_Text_Making

Brown, S. (2011). *Listening myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press

Campoy-Cubillo, M.C. & Querol-Julián, M. (2015). Assessing multimodal listening. In B. Crawford & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings: From research to teaching* (pp. 193-212). Routledge.

Chang, A. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*, 652-663.

Chang, A., & Millett, S. (2015). Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System, 53*, 91-102.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-170.

De Santiago Guervós, J., & Fernández González, J. (2017). *Fundamentos para la enseñanza del espalol como 2/L.* Madrid: Arcos.

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237-261. https://doi.org/10.2307/20445397

Donaghy, K. (2019, September 23). Advancing learning: the fifth skill –'viewing' [Blog entry]. MacMillanEnglish. Retrieved 21 April 2021, from https://www.macmillanenglish.com/us/blog-resources/article/advancing-learning-the-fifth-skill-viewing

Flowerdew, J., & Miller, L. (2005). *Second language listening*. New York: Columbia University Press.

Gil Fernández, J. (1990). *Los sonidos del language*. Madrid: Síntesis.

Gómez Serrano, M., et al. (2013). El oído como órgano receptor del sonido. In M. A. Penas Ibáñez (Ed.), *Panorama de la fonética española actual*. Madrid: Arcos.

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85-93.

Helgesen, M. (2003). Teaching listening. In D. Nunan (Ed.), *Practical English language teaching*. New York: McGraw-Hill.

Hsieh, Y. (2019). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer assisted language learning, 33*, 567-589. https://doi.org/10.1080/09588221.2019.1577898

Hulstijn, J. H. (1997). Mnemonic methods in foreign language vocabulary learning. In Coady & Huckin (Eds.), *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*: Routledge

Jewitt, C. (2013). Multimodal teaching and learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 4109-4114). Oxford: Blackwell Publishing Ltd.

Jewitt, C., Bezemer, J., & O'Halloran, K. L. (2016). *Introducing multimodality*. London & New York: Routledge.

Kartal, G., & Simsek, H. (2017). The effects of audiobooks on EFL students' listening comprehension. *The Reading Matrix, 17*(1), 112-123.

Kress, G. R., & Van Leeuwen, T. (2001). *Multimodal discourse: the modes and media of contemporary communication*. Oxford: Oxford University Press.

Kress, G. R., & Van Leeuwen, T. (2006). *Reading images: the grammar of visual design* (2nd ed.). London: Routledge.

Kress, G. R. (2010). *Multimodality. A social semiotic approach to contemporary communication*. London: Routledge.

Lee, J. A., & Van Patten, B. (2003). *Making communicative language teaching happen.* New York: McGraw Hill.

Legros, D., & Crinon, J. (2002). *Psychologie des apprentissages et multimedia.* Paris: Armand Colin.

Marrero, V. (2001). *Fonética perceptiva – Addenda*. Complementary material of the subject *Fonética y Fonología de la lengua Española*. Madrid, Departamento de Lengua Española y Lingüística General de la Facultad de Filología (UNED).

Martínez Lirola, M. Experiencia de enseñanza multimodal en una clase de idiomas. *Ensayos. Revista de la Facultad de Educación de Albacete*, *18*, 1-13. Retrieved from http://www.revista.uclm.es/index.php/ensayos

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 1-13.

Norte Fernández-Pacheco, N. (2016). Multimodal digital tools and EFL audio-visual comprehension: Students' attitudes towards vodcasts. *Language Value, 8*(1), 49-76.

Nunan, D. (2002). Listening in language learning. In Richards. J. C., & Renandya, W. A. (Eds.), *Methodology in language learning: an anthology of current practice* (pp. 238-241). Cambridge: Cambridge University Press.

Paivio, A. (2007). *Mind and its evolution: a dual coding theoretical approach.* Mahwah, NJ: Erlbaum.

Pérez Bernabeu, A. (2019). Comprensión auditiva y percepción multimodal: una nueva mirada a la comprensión de la oralidad desde la coherencia al paradigma comunicativo de enseñanza de idiomas. *Doblele, 5*(1), 47-65.

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon, 9*(5), 1-6. https://doi.org/10.1108/10748120110424816

Ramírez, D., & Alonso, I. (2007). Using digital stories to improve listening comprehension with Spansih young learners of English. *Language Learning & Technology, 11*(1), 87-101.

Rubin, J. (1995). The contribution of video to the development of competence in listening. In D. J. Mendelsonhn & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 151-165). San Diego, CA: Dominie Press.

Serani, A.; Behtash, E. Z., & Nezhad Arani, S. M. (2014). The effect of video-based tasks in listening comprehension of Iranian pre-intermediate EFL learners. *Gist Education and Learning Research Journal, 8*, 29-47. https://doi.org/10.26817/16925777.112

Smith, L., & Nelson, C. (1985). International intelligibility of English: directions and resources. *Englishes, 4*(3), 333-342. http://dx.doi.org/10.1111/j.1467-971X.1985.tb00423.x

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*, 661-699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Tavakoli, M., & Gerami, E. (2013). The effect of keyword and pictorial methods on EFL learners' vocabulary learning and retention. *Porta Linguarum, 19*, 299-316.

Unsworth, L. (2010). Resourcing multimodal literacy pedagogy. Toward a description of the meaning-making resources of language-image interaction. In T. Locke (Ed.), *Beyond the Grammar Wars* (pp. 276-293). London: Routledge.

Vandergrift, L. (2011). Listening: theory and practice in modern foreign language competence. Retrieved April 2, 2021, from Centre for languages, linguistics and area studies: www.llas.ac.uk/resources/gpg/67

Waring, R., & Nation, I. S. P. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Description, acquisition and pedagogy* (pp. 6-9). Cambridge: Cambridge University Press.

Wagner, E. (2007). Are they watching? An investigation on test-taker viewing behaviour during an L2 video listening test. *Language Learning & Technology, 11*(1), 67-86.

Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*(4),493-513.

Winiharti, M., & Herlina, C. (2017). Audio only or video? Multimodality for listening comprehension. *Advances in Social Science, Education and Humanities Research (ASSEHR), 82,* 189-192. https://doi.org/10.2991/conaplin-16.2017.41

Zoghbor, Z. (2016). A model for speech processing in second language listening activities. *English language teaching, 9*(2),13-19. http://dx.doi.org/10.5539/elt.v9n2p13

APPENDIX

1 Tool 1: materials

1a) Listening comprehension tasks (LO and LW)

Access the tasks in the following link:
https://drive.google.com/file/d/18IWDpXq3hKF3DgOFzKY-cpR_fIHnQDgY/view

1b) Answers of listening comprehension tasks (LO and LW)

Access the tasks in the following link:
https://drive.google.com/file/d/1CamgdZ8RuFzPMLfNExzeXQXnKgCyVzdI/view

1c) Pre-task for the two listenings

Access the tasks in the following link:
https://drive.google.com/file/d/15fdEXSI49T_Y9C_o2zScSgBGLCltsknH/view

2 Tool 2: questionnaire

Access the tasks in the following link:
https://docs.google.com/forms/d/e/1FAIpQLSdIryf01KIjeMoRRNNNYXmVAtx5FLCLY4Z
NNfdIybdroowknA/viewform?gxids=7628

3 Results tool 1 (tables)

3a) Total scores per subject in each format

	Subject	Scores LO	Scores LW	Difference between formats (in favour of LW)
LO + LW	1	6	9	3
	2	9	10	1
	3	2	7	5
	4	13	12	-1
	5	4	10	6
	6	2	11	9
	7	6	10	4
	8	8	11	3
	9	11	12	1
	10	12	13	1
	11	5	8	3
LW + LO	12	8	11	3
	13	12	12	0
	14	8	11	3
	15	10	12	2
	16	7	13	6
	17	10	12	2
	18	3	9	6
	19	8	11	3
	20	12	14	2
	21	10	12	2
	22	8	13	5
Overall average		7,909090909	11,04545455	3,136363636
Average group LO + LW		7,090909091	10,27272727	3,181818182
Average group LW + LO		8,727272727	11,81818182	3,090909091
Difference per group		1,636363636	1,545454545	0,09090909091

3b) Scores item by item

Performance LO format (item by item)

	Subject	It.1	It.2	It.3	It.4	It.5	It.6	It.7	It.8	It.9	It.10	It.11	It.12	It.13	It.14	Total (/14)
LO + LW	1	1	1	0	0	0	1	1	1	0	0	1	0	0	0	6
	2	1	1	0	0	1	1	1	0	0	0	1	1	1	1	9
	3	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2
	4	1	1	1	0	1	1	1	1	1	1	1	1	1	1	13
	5	1	1	1	0	0	0	1	0	0	0	0	0	0	0	4
	6	0	1	0	0	0	0	0	0	1	0	0	0	0	0	2
	7	1	1	0	0	1	0	1	0	0	0	1	0	1	0	6
	8	1	1	1	0	0	1	1	1	0	0	0	1	0	1	8
	9	1	1	1	1	1	1	1	1	0	0	1	1	0	1	11
	10	1	1	1	1	1	1	1	1	0	1	1	1	0	1	12
	11	1	1	0	0	0	0	0	0	0	0	1	1	0	1	5
LW + LO	12	1	1	1	0	1	0	0	0	0	0	1	1	1	1	8
	13	1	1	1	1	1	0	1	1	1	0	1	1	1	1	12
	14	1	1	1	0	1	1	0	1	1	0	1	0	0	0	8
	15	1	1	1	0	1	1	1	0	1	0	1	0	1	1	10
	16	1	1	1	0	1	0	1	0	0	1	1	0	0	0	7
	17	1	1	1	0	1	1	1	1	0	1	1	0	0	1	10
	18	1	1	0	0	1	0	0	0	0	0	0	0	0	0	3

	It.1	It.2	It.3	It.4	It.5	It.6	It.7	It.8	It.9	It.10	It.11	It.12	It.13	It.14	Total
19	1	1	1	0	1	0	0	1	1	1	0	0	0	1	8
20	1	1	1	0	1	1	1	1	1	0	1	1	1	1	12
21	1	1	1	0	1	1	1	1	0	0	1	1	0	1	10
22	1	1	1	0	1	0	0	1	1	0	1	1	0	0	8
AVERAGE	0,9090909091	1	0,68	0,1363636364	0,7272727273	0,5	0,6818181818	0,5454545455	0,3636363636	0,2272727273	0,7272727273	0,5	0,3181818182	0,5909090909	7,909090909

Performance LW format (item by item)

	Subject	It.1	It.2	It.3	It.4	It.5	It.6	It.7	It.8	It.9	It.10	It.11	It.12	It.13	It.14	Total (/14)
LO + LW	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	9
	2	1	1	1	1	0	0	1	1	0	1	1	1	1	0	10
	3	1	1	1	1	0	1	1	1	0	0	0	0	0	0	7
	4	1	1	1	1	1	0	1	1	1	1	1	1	1	0	12
	5	1	1	1	1	0	0	1	1	0	1	1	1	1	0	10
	6	1	1	1	0	0	0	1	1	1	1	1	1	1	1	11
	7	1	1	1	1	0	1	1	1	1	1	1	0	0	0	10
	8	1	1	1	1	1	0	0	0	1	1	1	1	1	1	11
	9	1	1	1	1	0	0	1	1	1	1	1	1	1	1	12
	10	1	1	1	1	0	1	1	1	1	1	1	1	1	1	13
	11	1	1	1	1	0	0	1	0	0	1	1	0	0	1	8
LW + LO	12	1	1	1	1	1	0	1	1	1	1	1	0	1	0	11
	13	1	1	1	1	0	1	1	1	1	1	1	1	1	0	12
	14	1	1	1	1	0	0	1	1	1	1	1	1	1	0	11
	15	1	1	1	1	0	1	1	1	1	1	1	1	1	0	12
	16	1	1	1	1	1	0	1	1	1	1	1	1	1	1	13
	17	1	1	1	1	1	0	1	1	1	1	1	1	1	0	12
	18	1	0	1	1	0	0	1	0	1	1	1	1	1	0	9
	19	1	1	1	1	1	1	1	0	0	1	1	1	1	0	11
	20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
	21	1	1	1	1	0	0	1	1	1	1	1	1	1	1	12
	22	1	1	1	1	0	1	1	1	1	1	1	1	1	1	13
AVERAGE		1	0,9545454545	1	0,9545454545	0,3181818182	0,3636363636	0,9545454545	0,8181818182	0,7272727273	0,9545454545	0,9545454545	0,7727272727	0,8636363636	0,4090909091	11,04545455

4   Results tool 2

4a) Exposition to formats

Subject	Watching films/series	Watching Youtube videos	Playing videogames	Listening to podcasts	Listening to music	Listening to the radio
1	1	2	3	1	4	1
2	3	3	5	1	5	3
3	2	3	1	3	5	2
4	4	4	4	3	5	3
5	3	3	2	1	5	1
6	3	4	1	1	5	3
7	1	2	1	1	3	1
8	5	4	2	1	4	1
9	3	4	5	1	5	1
10	1	5	5	2	5	1
11	4	2	1	1	5	2
12	4	1	1	3	5	2
13	3	2	1	1	4	1
14	3	4	2	2	5	1
15	5	2	2	1	4	5
16	5	2	2	1	5	2
17	2	2	1	1	4	1
18	2	3	1	3	4	2
19	5	5	3	2	5	2
20	2	3	2	1	3	1
21	3	4	4	2	4	4
22	3	3	4	1	4	1
AVERAGE	3,045454	3,045454	2,409090	1,545454	4,454545	1,863636
Standard deviation	1,290	1,0900	1,46901	0,80043	0,67098	1,12527

4b) Agreement with parameters

Subject	It. 1	It. 2	It. 3	It. 4	It. 5	It. 6 (value reversed)	It. 7 (value reversed)	It. 8	It. 9
1	5	2	4	5	5	5	4	3	5
2	4	2	2	4	5	5	5	5	5
3	4	3	3	5	5	4	4	4	4
4	4	2	4	3	4	3	4	3	4
5	4	3	3	5	4	4	5	3	4
6	4	3	3	3	5	4	3	4	3
7	4	1	4	4	5	5	5	5	5
8	5	2	4	3	5	5	4	5	5
9	4	2	2	4	5	4	4	1	4
10	4	2	4	4	4	4	3	4	4
11	3	2	4	4	2	4	4	4	4
12	5	2	2	3	4	4	4	4	4
13	4	3	3	2	5	3	2	2	4
14	5	2	5	4	5	5	5	4	5
15	4	3	3	3	5	4	4	4	5
16	5	2	4	4	5	5	4	5	5
17	4	3	3	3	5	5	4	4	5
18	4	3	4	4	4	4	4	4	4
19	5	2	3	2	5	4	4	4	5
20	2	1	1	4	2	5	5	2	4
21	5	3	3	2	4	4	4	4	4
22	4	2	4	4	5	5	5	5	5
AVERAGE	4,18	2,27	3,27	3,59	4,45	4,31	4,09	3,77	4,4
Standard deviation	0,73	0,63	0,93	0,90	0,91	0,64	0,75	1,06	0,59

4c) Parameters answer distribution

Item 1 (LW better sound quality)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	1	1	13	7

Item 2 (LW too long)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
2	12	8	0	0

Item 3 (LO too long)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
1	3	8	9	1

Item 4 (speakers spoke slower)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	3	6	10	3

Item 5 (More entertaining)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	2	0	6	14

Item 6 (easier to understand general info)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	0	2	11	9

Item 7 (more clear articulation)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	1	2	13	6

Item 8 (less anxious)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
1	2	3	11	5

Item 9 (easier retrieval)

Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
0	0	1	11	10