This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Alsalmi, Sami S

Title:
Acquisition of Implicit Knowledge of Second Language Syntax

The Effects of Input Modality and Working Memory

Acquisition of Implicit Knowledge of Second Language Syntax: The Effects of Input Modality and Working Memory

Sami Sulaiman Alsalmi

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the School of Education, Faculty of Social Sciences and Law

November 2019

73,596 words

Abstract

Typical research in second language (L2) implicit knowledge acquisition has centred on the role of learning conditions (e.g., explicit learning, focus on form, and oral output) in stimulating automatic processing in the L2 developing system. However, the process of automating L2 grammatical knowledge can be affected not only by learning mechanisms but also by the modality of stimulus presentation and working memory (WM) capacity. The present investigation aimed i) to examine whether encoding modality (auditory vs. visual) affects L2 implicit grammatical knowledge acquisition differently, and ii) to explore the modulating effects of WM capacity on the rate of emergence of L2 implicit knowledge in each encoding modality. In each of the two experiments, Experiment 1 with Chinese first-language speakers ($n = 77$) and Experiment 2 with Arabic first-language speakers ($n = 37$), participants were split into two groups, one of which was trained on three syntactic structures (tag questions, negative adverbs, and counterfactual conditional) under auditory exposure, whereas the other group was trained on the same structures but under visual exposure. The development of implicit grammatical knowledge, accrued from each input modality, was assessed via a pretest/posttest design using the measures of implicit knowledge (i.e., a timed grammaticality judgement task [timed GJT] and an elicited imitation task [EIT]). The WM capacity was measured using an operation-word-span (OSPAN) task. The results of both experiments demonstrated that auditory and visual exposure to L2 English syntax were found to result in successful emergence of implicit grammatical knowledge, as determined by both implicit knowledge measures: timed GJT and EIT. Furthermore, only in Experiment 1 was WM capacity found to correlate positively with automatized knowledge development accrued from auditory exposure on the GJT, but not on the EIT. However, the results demonstrated no significantly differentiable effects between visual and auditory interventions, but there was an observed trend towards visual modality advantage in grammatical knowledge automatization

in Experiment 1 only, in both timed GJT and EIT. Overall, the study suggests that modality effects may indeed be another relevant factor in stimulating the automatic processing in the L2 developing system. I argue, from a psycholinguistic view, that computational demands exerted by auditory modality on WM capacity are greater than those exerted by visual modality, resulting in detrimental effects on the development of automatic processing.

Keywords: *implicit grammatical knowledge, elicited imitation task (EIT), timed grammatical judgement task (timed GJT), working memory (WM), auditory modality, visual modality*

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Sami Sulaiman Alsalmi

DATE: 22/11/2019

Acknowledgement

After giving my thanks to God (Allah) for His assistance and the divine destiny by which I am being driven, I wish to acknowledge my deep gratitude to my parents, Sulaiman Alsulmi and Modi Alobaid for their extended support; they have remained waiting patiently over the past years for an answer to their repeated question: *When will you finish?*

This PhD was conducted under the supervision of Guoxing Yu and Nina Kazanina, both of whom have vastly contributed to the ideas and experimental design, and provided valuable support on the write-up of the chapters. Thank you very much. Furthermore, I truly owe a debt of gratitude to the School of Education which has prepared me with the prerequisite research courses prior to and during the research for this thesis. Without these courses: *introduction to statistics, multivariate statistics, psychophysiological methods in educational research; philosophy and research design in the social sciences; understanding educational research; introduction to qualitative research in social sciences,* the quality of my PhD would not have improved to such an extent.

I would like also to thank Angeline Barrett, the director of the PhD programme, for her support and continuous enquiries about my progress and if it was running smoothly.

I am very indebted to my wife, Manar Alshemali, for all her great support, who, indeed, undertook many family responsibilities although she was also busy with her English and MA studies. I would also like to thank my little sons, Sulaiman and Essam, who were with me throughout most of my PhD stay and underwent an expatriate life through no choice of their own. Maisam, my new daughter, might be the luckiest to come to this life without experiencing living away from the extended family. I must also thank my aunt, Miznah Alobaid, for her considerable attempts to make my family's life smoother, including looking after my children in many situations.

Finally, I would like to acknowledge that the images used in the experiment are retrieved from Google Images, but the exact references for some images are unavailable.

Table of Content

List of Tables

List of Figures

Chapter 1: Introduction

People invariably use implicit knowledge in daily activities and perceptions (J. N. Williams, 2009). The mother who drives her children to school every day subconsciously drives a manual car without the need to think about switching gears, while her conscious attention may be drawn to planning the day with her children. When people change radio stations, looking for stations in their mother tongue, they might promptly distinguish between various foreign languages used by stations that have normally been broadcast in their country, even though they have never had training in other languages. People speak their first language spontaneously—with their families at home or with colleagues at work—without much effort or concentration. Furthermore, they are able to distinguish anomalous use of language instantly, despite the fact that they might not be able to describe the underlying rules governing these structures. In a broad sense, **implicit knowledge** is knowledge that is deployed automatically in that people are generally unaware of possessing it (R. Ellis, 2009a; Rebuschat & Williams, 2013). Such knowledge "can be causally efficacious in the absence of awareness that this knowledge was acquired or that it is currently influencing processing, that is, in the absence of metaknowledge" (Cleeremans, Destrebecqz, & Boyer, 1998, p. 406).

The situation is drastically different for many second language (L2) learners. Many of them find themselves unable to speak their second language spontaneously, aware of the necessity to monitor their L2 regularities, specifically in situations that require prompt response such as a conversation over the phone. The need to monitor their grammar when speaking in a foreign language, or even prior to speaking, indicates that they have not acquired implicit knowledge of the necessary grammar, despite being familiar with the grammatical rules themselves (i.e., only **explicit knowledge** is acquired, R. Ellis, 2009a; J. N. Williams, 2009). This results in considerable effort in producing L2 and can, in turn, lead to frustration. In some instances, this might result in reluctance to speak, or fear when speaking

in the L2. Therefore, the overall aim of the present work was to understand how L2 grammatical knowledge can become automatized and deployed without much effort or explicit monitoring.

In this introductory chapter, four basic concepts underlying the current investigation will first be elucidated: implicit and explicit knowledge, measures of implicit grammatical knowledge, input modality, and working memory (WM) capacity. The nature of the research problem will be subsequently discussed, identifying the gaps the thesis aimed to address. Following the description of the research problem, the significance for the present study will be discussed from psycholinguistic and pedagogical perspectives, followed by the discussion of specific research questions. Finally, the chapter will provide an outline of the chapters intended to construct the argument of this thesis.

1.1 Basic concepts underlying the thesis

1.1.1 Constructs of implicit and explicit knowledge

The notion that L2 learners may possess two separate types of knowledge dates back to early theories of L2 acquisition. Krashen (1981), for instance, distinguished between an 'acquired system' and a 'learnt system'. The former results from subconscious learning processes (i.e., implicit learning) which arise when learners use language for communication. The latter, on the other hand, results from application of an attentional process (i.e., explicit learning) in an effort to comprehend and memorise grammatical regularities. The learnt system may work as an inspector of the acquired system (i.e., correctness of utterances). Krashen (1982, as cited in R. Ellis, 1994) later equated the acquisition versus learning distinction with implicit versus explicit knowledge.

It is implicit rather than explicit knowledge that underlies linguistic competence in native speakers, which is reflected in actual speech production and comprehension (R. Ellis,

2009a). This is the case regardless of whether linguistic competence is delineated as innate

(i.e., Universal Grammar) or emergent, as in connectionist theories (i.e., a neural network of

information nodes that a learner builds up from input over time, see Section 2.1 for more

information). However, the major challenge L2 acquisition researchers have met is how they

can distinguish whether an L2 learner has implicit knowledge (i.e., acquired) or explicit

knowledge (i.e., learnt) of some L2 grammatical feature. To demonstrate that implicit

knowledge or explicit knowledge exists, it is necessary, to first reveal that some *development*

in the learner's linguistic system has occurred (R. Ellis, 2015b); second, it is important to

understand how this L2 development is represented and processed in use in terms of implicit

and explicit knowledge. According to R. Ellis (2009a), L2 implicit knowledge is deployed

spontaneously where linguistic regularities are processed with automaticity. L2 explicit

knowledge, by contrast, is conscious and declarative, and it is available only through

controlled processing and intentional action, such as monitoring utterances. That is, any

language behaviour becoming available for use in automatic processing affords evidence of

the L2 learner's implicit knowledge (R. Ellis & Roever, 2018).

1.1.2   Measures of second language implicit knowledge

A number of research studies arising from factor analysis and correlational studies

have attempted to establish measures that tap specifically into implicit or explicit knowledge

(Bowles, 2011; R. Ellis, 2005; Erlam, 2006; J.-e. Kim & Nam, 2017; Loewen, 2009; Sarandi,

2015; Spada, Shiu, & Tomita, 2015; Suzuki & DeKeyser, 2015). These studies consistently

conclude that an elicited imitation task (EIT) and a timed grammaticality judgement task

(timed GJT) are better measures of L2 implicit knowledge, whereas an untimed GJT and a

metalinguistic knowledge test are better measures of L2 explicit knowledge. The most

dominant factor associated with the measures of implicit knowledge requires fluent and

speedy access to L2 knowledge where the likelihood of monitoring or pondering the

appropriate syntax, for example, is suppressed (i.e., accessing explicit knowledge, see more explanation about the understanding of automaticity under Section 2.5.3). Thus, both two tasks (i.e., EIT and timed GJT) involve time constraints where each possessing its own relative processing speed involved while automatically executing a task. The timed GJT, for example, measures the initial intuition about grammaticality where fast processing is the characteristic the timed GJT relies on in measuring implicit knowledge. The EIT, on the contrary, involves a deeper level of processing hinging on fluency in the auditory input and oral output as a dynamic, real-time entity. Hence, fast processing in the EIT is not the sole factor which reflects automaticity, rather it is the change in the way processing is executed (i.e., some form of restructuring, see Section 2.5 for a comprehension discussion about these two measures).

Following the development of the valid competence measures (i.e., EIT and timed GJT) to assess implicit grammatical knowledge, L2 acquisition researchers were in a position to scrutinise the learning mechanisms which might foster the acquisition of L2 implicit knowledge. However, such a one-dimensional focus (i.e., learning mechanisms) may not be sufficient to elucidate the acquisition success of implicit knowledge. That is, L2 learners' successful acquisition of implicit grammatical knowledge can be affected not only by learning conditions but also by the modality by which input is presented.

1.1.3   Input modality

The impact of input modality on L2 processes was of specific interest due to the hypothesis that auditory and visual inputs are represented and processed in two separate mechanisms, with each stream possessing its own properties (Penney, 1989). According to Anderson (1990), Danks (1990) and Rost (1990), the distinction between the two modalities is primarily manifested at the level of input access. That is, in auditory processing, to find word boundaries within a sentence the human brain must deal with prosodic signals (i.e.,

phonological recognition of words, stresses, and intonational patterns), which vary from speaker to speaker. Other idiosyncrasies of auditory language processing include the rapid speed of linguistic input (i.e., a continuous stream of sound is segmented throughout the time that new materials are perceived) and a high degree of transience (i.e., the input stream cannot be re-inspected). This means that a variety of cognitive functions must be at work in real time to parse the rapid linguistic input into its constituent units. This is unlike visual processing, where the processer parses the input in the absence of elements related to prosody; or in other words, representations of visual language are orthographically immutable. Compared with the auditory modality, visually presented language further allows for repetitive processing and permits the reader to control the rate of input (i.e., it allows for close analysis of text).

In light of the differences between the two modalities in terms of input access, the auditory modality appears to have distinctive idiosyncrasies which appears to require more substantial WM resources (including attentional resources) than visual modality.

1.1.4 Working memory capacity

Working memory (WM) capacity is delineated in terms of the ability to hold and manipulate information simultaneously as much as possible over a given temporal interval. This type of memory is considered the most crucial cognitive faculty in the human brain (Haberlandt, 1997), because it is not only limited to the storage of information per se (as is the case with short-term memory), but also associated with the processing and manipulation of information through functions such as thinking, reasoning, attention and comprehension.

In order to distinguish L2 learners with high WM capacity from those with low WM capacity in the efficiency of input processing, Juffs (2015) suggested that some processing demands Tmust be exerted on the parser to generate computational complexity. As such, in the case of the difference in processing between the two input modalities, the heavy demand

will be most likely to manifest in the processing in the auditory modality. This is because the processor must parse the rapid auditory input by recognising acoustic elements such as intonation and accentual system, and further linguistic elements like phonemes, syntax, and lexical items, resulting in more factors competing for the learner's limited WM resources. Therefore, the WM capacity is predicted to be a stronger factor in L2 acquisition in the auditory modality rather than the visual modality.

1.2   The research problem

Previous research investigating modality effects on L2 processes relied on data elicited from performance (i.e., not learning) where syntactic or comprehension processing was assessed (e.g., Murphy, 1997; Park, 2004; Wong, 2001). These studies found that the processing of linguistic information is more effective when such material is presented visually rather than auditorily. However, research investigating the effects of input modality on L2 learning itself is very limited (i.e., Sagarra & Abbuhl, 2013; Sydorenko, 2010); the existing results have been inconclusive and some revealed counter evidence (i.e., an advantageous bias towards auditory input). These results lead to a question of whether the effects of input modality on L2 performance are different from those on L2 acquisition. Therefore, the present study furthers research on this topic by scrutinizing whether encoding modality (auditory vs. visual) affects L2 implicit grammatical knowledge acquisition differently. However, what makes the present investigation different from the previous studies is that the language development, accrued from each modality intervention (auditory vs. visual input modalities), is assessed by the measures of implicit knowledge (i.e., an EIT and a timed GJT), following the assumption that such measures would provide more critical data regarding L2 acquisition. That is, they are more likely to elucidate whether the L2 developing system is predisposed to automating (i.e., successful emergence of implicit knowledge) via auditory or visual input based on language learning.

The present investigation further explores whether WM capacity predicts how this implicit knowledge is acquired differently with exposure to one of the two input modalities (auditory vs. visual). Previous research has provided evidence that WM plays a critical role in performance processes during text reading (Alptekin & Ercetin, 2010; Harrington & Sawyer, 1992; Jeon & Yamashita, 2014; Leeser, 2007), while sufficient evidence was not obtained for the interaction between WM capacity and L2 auditory performance (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012; Kormos & Safar, 2008; Miyake & Friedman, 1998). The results of these studies, however, contradict the general claim suggesting that it is auditory rather than visual input modality which exerts processing demands on WM capacity, resulting in a negative effect on L2 performance. Moreover, research investigating the relationship between WM capacity and the amount of learning accrued from one of the input modalities (auditory vs. visual) is very limited. Only one study (Sagarra & Abbuhl, 2013) to the best of the author's knowledge, has addressed this relationship and found results corroborating the general claims (i.e., it is the effect of auditory, not visual input, on the development of L2 learning which is mediated by WM capacity). In spite of the successful work in Sagarra and Abbuhl's (2013) study in terms of WM capacity, the substantial challenge remains regarding whether WM capacity still interacts with the L2 development, accrued from each input modality (auditory vs. visual), when this development is gauged through measures of implicit knowledge.

The lack of research that manipulated the modality of stimulus presentation and WM factors while measuring implicit knowledge is a noticeable oversight. In view of the fact that the two—modality and WM capacity—typically have influential values in both psychological and language acquisition research, this thesis offers a significant contribution to our further understanding of the L2 implicit knowledge acquisition. Because these aspects have not been

sufficiently described in the existing research, they are thus worthwhile issues for further empirical investigation.

1.3 The research significance

This study is necessary not only to demonstrate how the effect of a different intervention reveals different results but to contribute to an expanding psycholinguistic knowledge base on the acquisitional process of implicit grammatical knowledge. That is, the assessment of implicit knowledge development via an EIT and a timed GJT could make it more obvious whether the L2 developing system is predisposed to becoming automated (i.e., successful emergence of L2 implicit knowledge) in the case of auditory or visual input-based language learning. Similarly, the examination of the WM capacity provides a better understanding of the mediating variable that could increase or hinder the development of L2 implicit knowledge accrued from each modality. The project further contributes to a better understanding of i) the relationship between WM and automatic restoration required in the EIT and ii) the rapid response involved in the timed GJT.

From a pedagogical perspective, the findings of this study will be useful for language educators to better understand the extent to which auditory and visual modalities of learning L2 regularities can accelerate the automaticity of grammar use. The findings are further fundamental to teachers' pedagogical understanding because they can improve awareness of the cognitive operation of WM with students. Specifically, they will be able to understand which modality of grammar instruction induces learners with high WM capacity to exhibit greater benefits than those with low WM capacity.

1.4 Research questions

In light of the discussion above, two research questions have been put forward:

1. Which of the two input modalities (auditory vs. visual) results in more successful acquisition of implicit grammatical knowledge?

2. To what extent does WM capacity mediate the acquisition of implicit grammatical knowledge in the case of auditory versus visual input modality?

1.5   Brief description of this dissertation by chapters

This chapter briefly addressed the four basic theoretical concepts of the present work (i.e., implicit knowledge, measures of implicit knowledge, input modality, and working memory (WM) capacity), the research problem, the research significance, the research questions, and an outline of the methodology in an attempt to answer the research questions. The remainder of this thesis is organised in seven chapters.

Chapter Two will discuss in detail the theoretical and empirical backgrounds of the four topics that lay the foundation for the present work. The chapter will begin with a detailed account of the literature in attempt to provide a clear distinction between implicit and explicit knowledge in second language research. This will lead to a discussion on how each type of knowledge is measured, addressing a comprehensive review of the previous major studies that provided evidence establishing the validity of a battery of tests hypothesised to tap into independent measures of implicit and explicit knowledge. The discussion will focus on the timed GJT and EIT (the measures of the L2 implicit knowledge) because it is the development of the L2 implicit knowledge which the present investigation examines. Echoing the calls initiated by innatist and connectionist theories of language learning to scrutinise the mechanisms which foster the L2 implicit knowledge, the chapter will then address whether the modality of input presentation (auditory vs. visual) possibly affects L2 acquisition differently. More relevant to the theme of modality's effects on L2 performance, the chapter will depict the role of WM capacity in L2 language phenomena and addresses how the effects of modality (auditory vs. visual input) on L2 performance, including the

development of L2 learning, are mediated by WM capacity. The end of Chapter Two will include a brief theoretical framework of the current research, highlighting the main findings of previous studies and emphasising the implications of those for the present investigation in terms of research design.

Chapter Three will provide a succinct summary of the research rationale together with the hypotheses and research questions. Next, the chapter will provide a comprehensive description of the methodology used by Experiments 1 and 2 (i.e., the same methodology was employed in both experiments with a difference in the first language of participants in each experiment). The methodology section will provide information about the target participants and then address the properties of the three stimuli structures (i.e., TQ, NA, and CON) separately in terms of five dimensions: processing, regularity, frequency, perceptual salience, and functional value, for the purpose of determining the level of difficulty for each structure. Following this description, the methodology section will provide a detailed description of the data collection, including the tasks used in the training stage (i.e., focus on form and deductive learning tasks) and testing stages (timed GJT and EIT 'as measures of L2 implicit knowledge', OSPAN 'as a measure of WM capacity', and digit span [DS] 'as a measure of phonological short-term memory').

Chapter Four will present Experiment 1, in which the participants' first language was Chinese. Experiment 1 sought to investigate the above-mentioned research questions. The chapter on Experiment 1will provide a detailed statistical analysis of the data obtained from the experiment conducted to test the hypotheses of the study. Finally, the chapter will end with a section discussing how the answers to the research questions are supported by the results of previous empirical studies and fit the existing body of knowledge about each research question.

Chapter Five will present Experiment 2, which followed the same training and testing procedures as Experiment 1. However, Experiment 2 differed from Experiment 1 in that the target participants were Arabic learners of English. The overall aim of Experiment 2 was to verify the results of Experiment 1. Similar to Experiment 1, the chapter on Experiment 2 will provide a detailed statistical analysis of the data and will end with a discussion section.

Chapter Six will provide a general discussion of the present study. The chapter will be developed using a summary of the study's motivation, a summary of the results, and consideration of the outcomes of Experiments 1 and 2. Next, the focus of the chapter will shift to the performance difference between timed GJT and EIT in the present investigation, addressing the underlying variables operationalised by both measures.

Chapter seven will conclude with the limitations, and suggestions for future research.

Chapter 2: Literature review

The second chapter of this thesis provides the theoretical background to the present work in relation to the development of L2 implicit grammatical knowledge, input modality, and working memory (WM) capacity as a starting point for the current investigation. The chapter first introduces the constructs of implicit knowledge and how it is distinct from explicit knowledge. Following this description, the chapter discusses the relative effects of implicit, incidental, and explicit learning on L2 acquisition. Next, the discussion turns to the measures of implicit grammatical knowledge, namely, the elicited imitation task (EIT) and grammaticality judgement task (GJT), and how these measures are distinct from those used for measuring explicit grammatical knowledge. Then, the chapter discusses whether output production practice plays a facilitative role in the success of L2 acquisitional process. Following this, the focus of the chapter shifts to clarify the theoretical and empirical bases on the distinctive processing between auditory and visual modalities and how the modality of input affects the acquisition of L2 knowledge. Moving on from input modalities, the chapter discusses the nature and constructs of WM and how this type of memory plays a role in L2 acquisition. Finally, the chapter provides a brief theoretical framework of the current research, highlighting the main findings of the reviewed studies and emphasising the implications for the present investigation in terms of research design.

2.1 Constructs of L2 implicit knowledge

Prior to elucidating the nature of implicit linguistic knowledge and how it is distinct from explicit knowledge, it is important to clarify what 'linguistic competence' implies. Two competing theories delineate the concept of linguistic competence (R. Ellis, 2009a). The first, following Chomsky's view (1965, 1982, 1986), claims that human beings are biologically endowed with an innate device mechanism (i.e., innate Universal Grammar), which functions as a guiding force for language acquisition. The assumption underlying Universal Grammar

indicates that language comprises a set of abstract principles that describe the central grammars of all natural languages. That is, language develops as a result of obtaining input from the environment with the support of an innate faculty. The second, drawing on connectionist theories of language learning, views linguistic competence as an elaborate, interlinked network of nodes that a human being constructs from input over time. That is, language learning, based on the connectionist view, arises neither from a deductive nor an inductive analysis of linguistic input, nor from the use of an innate language mechanism, but instead, it occurs built on the extraction of regularities from the input. In other words, it is the ability of the brain to search for units and nodes and make connections between them (N.Elli, 2003; cited in Mystkowska-Wiertelak, A. & Pawlak, M., 2012). These connections could be strengthened or weakened by repeated activation or inconsistent activation, respectively. Complex networks are developed as new relations are grown and new links are generated between larger and larger units (see Gass & Selinker, 2008). R. Ellis (2009a) claimed that although the two theories (i.e., Universal Grammar and connectionist theories) appear oppositional, they share the notion that linguistic competence primarily encompasses the implicit knowledge. Accordingly, the two theories constantly deliberate over investigating mechanisms of learning leading to the promotion of implicit linguistic knowledge because the development of such knowledge is the goal of their theory.

Admittedly, the definition of implicit and explicit knowledge is somewhat ambiguous (Robinson, Mackey, Gass, & Schmidt, 2013). In a broad sense, most researchers into L2 acquisition have been unanimous in their view that implicit knowledge fundamentally lies in automaticity. However, researchers in the psychology field consider a lack of awareness as indicative of implicitness. Both concepts are discussed in detail below.

2.1.1   Implicit knowledge defined as automaticity

Although no well-defined construct distinguishes automatic processing from non-automatic (or controlled) processing, L2 acquisition researchers have viewed rapid language processing with little or no conscious effort as a prevailing diagnostic tool that reflects acquisition rather than learning (Segalowitz, 2003; J. N. Williams, 2009). However, Segalowitz (2003) argued that, although it is reasonable to believe a mechanism functions more quickly once it becomes automatic, this does not necessarily indicate that it is entirely automatic. If the nature of automaticity is restricted merely to rapid processing, it is easier to establish a speed threshold in milliseconds that distinguishes automatic processing from non-automatic processing (Segalowitz, 2010, as cited in Segalowitz, 2013). Therefore, Segalowitz (2003) described automaticity in various dimensions, each of which referred to the type of processing used when automatically executing a task. Of these dimensions, those most associated with L2 acquisition include processing speed, ballistic processing (i.e., a process that cannot be disrupted by other ongoing processes once it begins), load-independent processing (i.e., automatic processing that functions regardless of the amount of information that must be processed), effortless processing (i.e., automaticity that does not require further effort and attention), and unconscious processing (i.e., automaticity that relies on priming effects developed during training).

None of these dimensions appears to function without considering processing speed. However, each dimension has its own relative processing speed standard that depends completely on the specific angle of automaticity in which the experimenter investigates in an experiment. This principle is quite similar to professionalism in a football game: the player must display certain characteristics to be professional, such as passing the ball, stealing the ball from an opponent, and defending and shielding it. All these features involve speed, but each has its own speed standard and associated skills, training, and assessments. Accordingly,

it is sensible to argue that processing speed, in a broad sense, is fundamental to automaticity. Segalowitz (2010, p. 79) claimed that the definitions of automaticity (i.e., the dimensions mentioned above) usually involve rapid processing, but the "speed of operation of the process cannot be the sole justification for calling it automatic." See Section 2.5.3 for an example of why rapid processing cannot be a mere synonym for automaticity. Therefore, the operational sense of automaticity used in the current study refers to fluent and speedy access to L2 knowledge in which the likelihood of monitoring the appropriate syntax, for example, is suppressed. How this access is measured will be discussed in Section 2.5.

2.1.2 Implicit knowledge defined as lack of awareness

From a cognitive psychology perspective, lack of awareness is the most prevalent basis of automaticity (J. N. Williams, 2009). It postulates that implicit knowledge also entails unconscious awareness where a language user is unconsciously aware of the underlying regularities governing linguistic structure. This is in contrast to explicit knowledge, which requires conscious knowledge of those regularities (R. Ellis, 2009a). The following example illustrates the two types:

** The new laptop that I bought **it** yesterday stopped working.*

A language user with implicit knowledge is intuitively aware that the sentence above is ungrammatical and could find its deviating element, but they may be unaware of the rule underlying the deviance. However, a user with explicit knowledge can verbalize that rule: the object of the verb is moved to sentence-initial position, and the original position after the verb must be omitted (i.e., a relative clause).

However, more importantly, language users could have both implicit and explicit knowledge of the same linguistic structure. That is, they may be able to use the structure correctly and fluently in their oral production (e.g., counterfactual conditionals: *If you had not passed the driving test, you would not have had a driving licence*) and simultaneously possess

conscious knowledge of the syntax guiding the structure: the past perfect tense is used following 'if' in the dependent clause, and 'would + have' followed by a verb in the past participle is used in the main clause.

This combined knowledge results from two sources: either the language users, whose L1 was acquired implicitly, have been exposed to the metalinguistic explanation (i.e., explicit learning) of their implicit language knowledge (i.e., native language). or implicit knowledge might have been derived from explicit knowledge when the L2 learners were exposed to meaningful, copious practice. This leads us to a crucial issue which has been a long-term focus in the field of L2 acquisition research: the role of implicit vs. explicit learning in the success of L2 acquisition.

2.2   Implicit and explicit learning: theoretical issues

Explicit and implicit learning are typically viewed as two oppositional conditions, and both have been a long-term focus in the L2 literature. Theoretically, the distinction between implicit and explicit learning can be basically determined by awareness of what is being learned and the intentionality of learning (R. Ellis, 2009a; J. N. Williams, 2009).

2.2.1   Implicit and incidental learning

N. Ellis (1994, p. 1) defined implicit learning as the "acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operations". Accordingly, implicit learning can be referred to as a manipulated learning condition in which participants learn grammatical structures in the absence of intentionality and awareness of the structure to be learnt at the point of learning (J. N. Williams, 2009). A realistic example of implicit learning is a child's first language acquisition. Especially prior to starting formal schooling (at approximately the age of six years old), children acquire complex linguistic knowledge spontaneously,

unintentionally, and implicitly without the ability to describe that linguistic knowledge (see Hulstijn, 2015, p. for a review). Thus, ever greater strides in language acquisition are noticeable in children, and their knowledge eventually outstrips that of the adults. An adult's learning of a second language, conversely, typically involves some degree of formal instruction and error correction (N. Ellis, 2011; Ioup, 1995, i.e., explicit learning). According to Bley-Vroman's (1989) Fundamental Difference Hypothesis, adult learners are unable to access their innate faculty of language acquisition (if one maintains Chomsky's perspective), and they have to have recourse to using problem-solving skills to learn a second language. Thus, they specifically rely heavily on controlled processing (i.e., requiring more attentional resources, see section 2.8.3 for more discussion). This contrasts with children's language learning in that a lack of conscious attention to linguistic structure is a unique idiosyncrasy of their learning, and, as a result, their intuitive knowledge of their grammar correctness develops.

In light of the L2 adult learning discussed above, within adult L2 acquisition-based research, learning without planning or intending to can be methodologically feasible. For instance, participants might learn grammatical structure in the course of a meaning-focused activity (i.e., incidental learning, see Section 2.6.1 for more discussion). However, there is debate as to whether any learning is likely to occur without some degree of awareness. This raises a crucial inquiry regarding the exact meaning of 'awareness'. Schmidt (1990, 2001), who suggested the 'Noticing Hypothesis', distinguished between two types of awareness: awareness as noticing (involving perception) and metalinguistic awareness (requiring understanding). In the former, conscious attention to 'surface elements' is involved, whereas in the latter awareness of the underlying syntax that guides particular linguistics is involved. The hypothesis suggests that as manipulation of noticing (i.e., the target structure in the input is emphasised in some way: through glossing, underlining, intonation, and so on) involves at

least some degree of awareness, there is no such thing as a completely implicit learning condition. Thus, implicit learning can be defined, as Schmidt (2001) suggested, as learning without any metalinguistic awareness.

The conceptual and methodological complexity of implicit learning have led to it being distinguished from incidental learning in the field of L2 learning research. While implicit learning refers to learning that proceeds without any awareness of what is being learned and without the intention to learn it, the incidental case refers to knowledge that "is acquired in the absence of instructions or intention to learn, but the subject retains conscious access to that knowledge" (Pervin & John, 1999, p. 587). That is, both types of learning (implicit and incidental) are unintended (or unplanned), but the difference lies in consciousness; that is, implicit learning is unconscious (i.e., being unaware of acquired knowledge) while incidental learning is conscious (i.e., being aware of knowledge being learnt). Rebuschat and Williams (2013) claimed that if exposure in an experiment, which has been planned to provide implicit learning, resulted in conscious (explicit) knowledge (i.e., participants were able to figure out the rule system although they were not informed about its existence), the learning process, in this case, would usually be considered as incidental and not as implicit learning. This simultaneously applies to those experiments which do not include a measure of awareness in the testing stage.

2.2.2 Explicit learning

While implicit learning involves the absence of intentionality and awareness of the structure to be learnt at the point of learning, explicit learning requires an intention to learn and the use of conscious knowledge at the point of learning. That is, participants are aware that they are supposed to learn something, and they are usually informed that there will be a test on what they have learned. As implicit learning indicates a lack of awareness of the learning process, it is assumed that the learning proceeds with a minimum of attentional

resources. However, in explicit learning, the metalinguistic knowledge is verbalised in the pattern of rules and, hence, more processing demands are placed on WM capacity (Y. Wang, 2017).

As the mechanisms involved in first language acquisition, which occurs implicitly, are not sufficient for L2 acquisition due largely to learnt attention and L1 transfer, the success of L2 acquisition needs to conquer the processing habits of the first language by deploying supplementary resources of explicit/intentional learning. N. Ellis (2002, p. 145) argued:

> Learned attention limits the potential of implicit learning, and that is why explicit learning is necessary in L2 acquisition … Language acquisition can be speeded by explicit instruction. The last 20 years of empirical investigations into the effectiveness of L2 instruction demonstrate that focused L2 instruction results in large target-oriented gains, that explicit types of instruction are more effective than implicit types, and that the effectiveness of L2 instruction is durable.

N. Ellis (2011), therefore, claimed that adult language acquisition usually involves explicit learning. Accordingly, the explicit learning process can be indicated as either inductive or deductive. According to Decco (1996, p. 96), deductive learning is a "process that goes from the general to the specific, from consciously formulated rules to the application in language use". To put it differently, metalinguistic knowledge (i.e., an explanation) of a grammatical rule is provided first, followed by examples and practice. Inductive learning, by contrast, refers to a process that goes from the specific to the general (i.e., forming generalisations) on the basis of authentic language containing examples of the rule where learners are then asked to discover the rule from those prototypical examples. In this regard, both types of deductive and inductive learning are considered explicit "because there is either explicit rule explanation or guided instruction for students to attend consciously to particular forms" (Y. Wang, 2017, p. 45).

2.3   The interface issue

The distinction between implicit/explicit knowledge is arguably less complicated than the implicit/explicit learning distinction (R. Ellis, 2015a). However, these distinctions raise three crucial questions: Can adult L2 learners acquire a foreign language in a complete implicit context as children do in learning their first language? Can explicit knowledge be converted into or facilitate the acquisition of implicit knowledge? Does explicit instruction result in the acquisition of implicit as well as explicit knowledge? (R. Ellis, 2009a). The understanding of relationships between implicit and explicit knowledge influenced significant discussions and debates among L2 researchers in the 1980s, generating three different positions. The first, which Krashen (1981, 1982) strongly maintained, was referred to as the non-interface position, which suggests that implicit and explicit knowledge are dissociated and distinct; they are located and accessed by different areas of the brain. Accordingly, implicit language knowledge is acquired only via exposure and cannot be explicitly taught.

The strong interface position, which was advanced by Sharwood Smith (1981) and was subsequently supported by DeKeyser (2003), claims that explicit knowledge can represent a base for practice. The resulting proceduralization eventually automatizes explicit knowledge. This automatized knowledge is reflected in the rapid online processing and becomes available through the automaticity and consistency of L2 learners' speech. Like implicit knowledge, automatized knowledge is accessed without or with very limited attentional actions regarding linguistic form. However, Dekeyser (2003) argued that this automatized knowledge learned explicitly should be distinct from knowledge learnt implicitly. In this sense, three types of knowledge can be built: explicit knowledge, automatized explicit knowledge, and implicit knowledge. Dekeyser (2003) argued that automatized knowledge can be deemed to be 'functionally equivalent' to implicit knowledge.

Finally, the weak interface position suggests that explicit knowledge can become implicit knowledge in an indirect manner. It postulates that this conversion develops only under specific constraints. According to R. Ellis (2009a), the first version posits that the transfer of explicit to implicit knowledge takes place if the target linguistic structure meets the learners' developmental readiness. The second version, theorized by N. Ellis (1994) and Paradis (1994), suggests that explicit knowledge can facilitate the development of implicit knowledge by "alerting learners to what they need to attend to when processing input" (Robinson et al., 2013, p. 252). MacWhinney (1997) supported this theory, finding that learning with some clues and hints produces better gains than learning without them. The third version of this position refers to the notion that learners can use explicit knowledge to generate output, and this output can work as 'auto-input' to the implicit system(Schmidt & Frota, 1986; Sharwood-Smith, 1981, as cited in R. Ellis, 2009a).

Each of the three distinct positions reviewed above has its advocates, and the concept has been a broadly controversial topic in L2 acquisition research. However, R. Ellis (2009a) argued the issue of the interface between these positions has not been subjected to an experimental scrutiny, which he claimed was due to the absence of consensus in terms of robust instruments for examining whether a learning effect occurred as a result of instruction or exposure composed of one type of knowledge (implicit or explicit), or some combination of the two. Y. Wang (2017), on the other hand, claimed that most recent studies tend to support the indirect, weak interface position; however, more examination and evidence is required to support this claim.

Even though the topic of the interface has been substantially disputed and no clear conclusions have been drawn from the discussions, R. Ellis (2015a) strongly argued that this should not necessarily be the case because implicit/explicit knowledge and implicit/explicit learning are associated. More specifically, implicit learning could result in explicit

knowledge, as a learner may become spontaneously aware of linguistic rules in the input (J. N. Williams, 2009) or when a learner consciously processes a chunk of knowledge that was acquired implicitly and then builds up an explicit generalisation about its linguistic form (R. Ellis, 2015a). This is specifically manifested in adult language learners in contexts where the learnt language is considered as an object rather than as a communicative tool. Explicit learning may also result in implicit knowledge. This can occur when a pupil explicitly learns a rule and then is exposed to copious and repeated practice; explicit knowledge becomes automatized and may come to affect the L2 developing system without the pupil's awareness (see R. Ellis, 2015a for a review). N. Ellis (2011) posited that the human mind is 'collaborative' in the sense that explicit knowledge continuously interfaces with the processes required in implicit learning. The following figure reveals the hypothesised relationships between implicit and explicit learning and implicit and explicit knowledge.

Implicit learning → Implicit knowledge

Explicit learning → Explicit knowledge
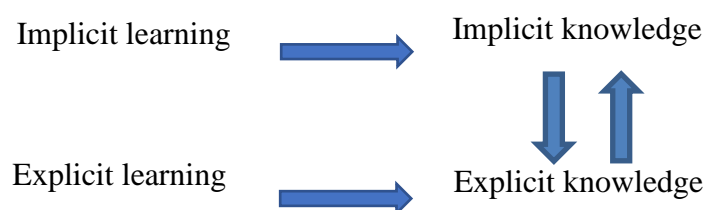
Figure 1. The relationships between implicit/explicit learning and implicit/explicit knowledge (taken from R. Ellis, 2015a).

As shown in Figure 1, the instruction that is employed for either type of learning might lead to either type of knowledge. This provides a clear indication that the learning experience is the impetus for establishing the type of knowledge acquired (R. Ellis, 2015a).

To conclude the theoretical discussion, the distinction between implicit and explicit learning can be basically determined by an awareness of what is being learned and the intentionality of learning. Language users could have both implicit and explicit knowledge of

the same linguistic structure. The dominant concept which guides the understanding of implicit knowledge in the L2 research is whether grammatical knowledge is available for use in automatic processing, which is reflected in fast processing. This can be supported by the claim by R. Ellis & Roever (2018, p. 2) that "automatic processing is a defining characteristic of implicit knowledge and, therefore, that any language behaviour involving automatic processing affords evidence of the learner's implicit knowledge". This is irrespective of the distinct theoretical views regarding whether explicit knowledge is proceduralized to become automatized knowledge or 'functionally equivalent' to implicit knowledge. The interface between these views should not necessarily be the case because either type of learning (i.e., implicit and explicit) is assumed to result in either type of knowledge (i.e., implicit and explicit).

2.4   Implicit and explicit learning: empirical studies

A wide range of experimental and quasi-experimental studies have been carried out to obtain insights into the impact of implicit /explicit learning upon the L2 acquisitional process and outcome (see J. M. Norris & Ortega, 2000; Spada & Tomita, 2010; J. Williams & Doughty, 1998, for a review). In implicit learning research, a number of research attempts have provided some empirical evidence that L2 learners can learn some L2 features without any awareness of the regularity being learned at the point of learning (i.e., Chen et al., 2011; Rebuschat & Williams, 2012; J. N. Williams, 2005).

J. N. Williams' (2005) study, which was the first to challenge Schmidt's Noticing Hypothesis, investigated the learning of form–meaning connections under conditions where the relevant forms were noticed, but the critical aspects of meaning were not. The participants (n = 41) were exposed to a semiartificial language consisting of four artificial determiners: *gi, ro, ul, and ne*. At the beginning, they were instructed that these determiners functioned like English determiners and coded distance where *gi* and *ro* were employed to express near

objects, while *ul* and *ne* were employed to indicate far objects. However, the participants were not informed whether the choice of determiner also depended on the animacy of the noun. The participants were trained in the distinction between near and far objects through sentences presented in an auditory mode (e.g., *the children threw sticks at <u>ul</u> monkey in the tree* or *I spent an hour polishing <u>ro</u> table before the dinner party*). The testing stage, in which they were not also informed that there would be a test of learning, consisted of two parts: the first part included a two-alternative, forced language choice and retrospective verbal reports that served as a measure of awareness. The results of the first part revealed that 33 out of the 41 participants expressed unawareness of the learnt language. The second part of the testing required unaware participants to search for a rule that determined the use of the target linguistic feature. According to the results, 11 out of the 33 participants were aware of the linguistic regularity that governed the structure. These findings were interpreted as evidence that the participants could acquire form-meaning mappings without becoming aware of the targeted linguistic system.

Although J. N. Williams' (2005) study was able to provide some evidence that adult language learners can acquire novel form–meaning mappings implicitly (i.e., unconscious knowledge was acquired where participants were generally unaware of possessing, Rebuschat & Williams, 2013, for more a discussion), Kerz, Wiechmann, and Riedel (2017) argued that other subsequent extensions of Williams' study could not replicate the finding that learning without awareness is possible (e.g., Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010). However, in a study conducted by Kerz, Wiechmann, and Riedel (2017), which is built on Williams' (2005) study, it was found that L2 knowledge can be acquired implicitly outside the laboratory settings. More support for learning without awareness is also derived from studies examining the other aspects of L2 learning, such as the acquisition of L2 morphosyntax (Rogers, Revesz, & Rebuschat, 2015) and word order (Rebuschat & Williams,

2012; J. N. Williams & Kuribara, 2008). In 'experiment 2' of Rebuschat and Williams' (2012) study, for example, a total of 28 native speakers of English were randomly assigned to experimental and control groups in which the experimental participants were exposed to a semi-artificial language that consisted of English words and German syntax as in the following the example:

English: *Yesterday John has bought the newspaper in the supermarket.*

German: *Gestern hat John die Zeitung im Supermarkt gekauft.*

Stimulus: Yesterday has John the newspaper in the supermarket bought.

Participants were trained in the semi-artificial regularities by way of untimed elicited oral imitations and plausibility judgments. That is, participants were auditorily exposed to sentences on an item by item basis and required to repeat each sentence after a delayed prompt, and had to judge whether the sentence they heard was semantically plausible or not. The testing stage included three tasks: i) untimed grammaticality judgement task (untimed GJT), ii) confidence ratings, and iii) source attributions. The first was used as a measure of learning while the second required participants to indicate their levels of confidence by selecting one of three response options: guess, somewhat confident, or very confident. In the case of the source attributions, the third testing task, participants were asked to indicate what the basis of their grammaticality judgment was by selecting one of four response options: guess, intuition, memory, and rule knowledge. The results provide some evidence that incidental exposure to L2 syntax can result in unconscious knowledge. This means that at least some of the learning was implicit. The analysis of confidence ratings and source attributions revealed that, although the participants were aware that they had acquired knowledge, they were at least partially unaware of what knowledge they had acquired. Participants ascribing their judgement to 'intuition' in the source attributions performed significantly above chance; put differently, unconscious syntactic knowledge was acquired.

Conscious, but unverbalizable knowledge obviously pertained to improved performance in the GJT.

Overall, although some studies were able to provide evidence that L2 adult learners can learn regularities in the absence of awareness at the point of learning, it appears that operationalising such learning is not straightforward and cannot be guaranteed to take place successfully. In addition, and more importantly, participants' prior knowledge must be completely controlled to provide exclusively implicit learning. Thus, it can be noted that previous studies (e.g., Rebuschat & Williams, 2012; J. N. Williams, 2005) employed an artificial or semi-artificial language to control such prior knowledge.

Accordingly, in most empirical L2 language acquisition which employed natural language (e.g., English) to generate the stimulus material, implicit learning was generally operationalised in the absence of regularity instruction and guidance to attend to a specific linguistic feature (see J. M. Norris & Ortega, 2000, for a comprehensive review). Explicit learning, by contrast, was operationalised by providing some sort of regularity presentation or explicit direction to promote metalinguistic consciousness of the regularity. However, it is worth noting that the operationalisation of the two types of learning is often based on the researchers' perspective in navigating their studies. That is, any learning/instruction type can be operationalised "as being more or less 'explicit' in an implicit and explicit continuum" (Norris and Ortega, 2000, as cited in Y. Wang, 2017, p. 36).

Robinson (1995), for instance, investigated the relative difference between four learning conditions: implicit, incidental, rule-search, and instructed learning. A total of 104 L2 learners were randomly assigned to one of the four groups. Participants in both the *implicit* condition (n = 26) and *incidental* condition (n = 26) were incidentally exposed to the L2 rules, i.e., they were neither informed that they would be exposed to learning nor that there would be an assessment of learning afterward. However, the basic distinction between

the two learning conditions lies in the allocation of focus: Participants in the implicit learning condition were required to allocate their focus to the ordering of the words in the stimulus sentences, while participants in the incidental learning condition were required to answer comprehension questions, and hence, they had to focus on the meaning of the stimuli. For the rule-search and the instructed conditions, the participants in the former condition were instructed to discover the L2 rules while being exposed to the training set, while participants in the latter condition received metalinguistic explanations of the rules plus rule-oriented training. The testing stage in all conditions included a GJT as a measure of what had been learnt and retrospective verbal reports as a measure of consciousness (i.e., how conscious the acquired knowledge was). The analysis of data obtained from the GJT indicated that participants in the instructed learning condition group scored highest, followed by the incidental group, then the rule-search group; the implicit group improved the least. The analyses of the verbal reports revealed that many participants in all conditions noticed rules during the process of learning and were able to report them when prompted to do so. This observation indicates that participants can acquire L2 syntax incidentally (as evidenced by the performance in the implicit and incidental groups), but the retrospective verbal reports revealed no evidence that training led to unconscious knowledge (i.e., they were aware of possessing).

In the same vein, Thuy-Minh, Hanh, and Pham (2012) conducted a quasi-experimental study to investigate the relative effects of implicit and explicit form-focused instruction on the acquisition of the speech act set of constructive criticism in English. A total of 69 L2 learners were randomly assigned to one of three groups: an implicit group (n = 19), an explicit group (n = 29), or a control group (n = 22). The explicit participants performed consciousness-raising activities by receiving an explicit metapragmatic explanation and correction regarding errors of forms and meanings. The implicit participants, on the other

hand, received exposure to enriched target pragmatic input via input enhancement and recast activities. The data were collected by means of three instruments: an 8-item written discourse completion task, a 6-item oral role play, and oral peer feedback on actual written works. The results revealed that both treatment groups significantly improved in the immediate and delayed posttest over the pretest, outperforming the control group. More importantly, the performance in the explicit group improved significantly more than the performance in the implicit group on all three measures. However, the study operationalised the implicit treatment by eliminating the language explanation, and the structures were enhanced for the participants. Furthermore, the testing stage did not include a measure of awareness (e.g., retrospective verbal reports) to determine what kind of knowledge was acquired: conscious or unconscious.

In Norris and Ortega (2000), an extensive meta-analysis of 49 studies was conducted to investigate the relative effects of explicit and implicit learning on the L2 acquisition based on the scale of Cohen's effect size (i.e., sizes larger than 0.8 is considered as *large*, sizes between 0.5 and 0.8 as *medium,* and sizes between 0.2 and 0.5 as *small*.) The authors classified the measures of language development into four types: metalinguistic judgment, selected responses, constrained constructed responses, and free constructed responses. Accordingly, the results revealed that the effect size for implicit learning is 0.54 whereas that for explicit instruction is 1.13, demonstrating that there is an advantage to explicit learning. In another study conducted by Spada and Tomita (2010), an extensive meta-analysis of 41 studies was conducted to investigate the relative effects of explicit and implicit learning on the acquisition of simple and complex grammatical features in English. Complexity had been described in terms of the number of transformations required to arrive at the target form. For instance, '*Wh*-question as object of preposition' as in *Whom did you talk to?* is distinguished as a complex feature, as seven transformations must be performed in order to arrive at the

sample sentence. In contrast, the past tense of regular verbs is described as simple because only one criterion (i.e., annexing the 'ed' inflection) must be met in order to arrive at the target form. The study relied on the same measures as were used in Norris and Ortega's (2000) study and divided them into two types: controlled and free constructed tasks. The former included metalinguistic judgments, selected responses, and constrained constructed responses, which were assumed to measure explicit knowledge while the latter included the free constructed response, which was assumed to measure implicit knowledge. The results revealed larger effect sizes for explicit over implicit treatment for simple and complex features. The findings further indicated that explicit learning plays a positive role in learners' controlled knowledge and spontaneous use of complex and simple structures.

Given the results of the studies above, a conclusion can be drawn that explicit learning is more effective than implicit learning for L2 language development. However, great caution should be exercised in interpreting the results in Norris and Ortega's (2000) research because most of the measures in the studies Norris and Ortega examined were more likely to tap into explicit knowledge rather than implicit knowledge. Doughty (2003, p. 273) argued that these tests, except for free constructed responses, "merely require knowledge of language as an object". This similarly applied to the studies investigated in Spada and Tomita's (2010) meta-analysis where it was unknown to what extent the free constructed response (e.g., a written composition or an oral narrative) does not permit learners to use their explicit knowledge.

2.5   Measures of L2 implicit knowledge

The necessity for enhanced valid competency measures to accurately assess implicit knowledge has recently occupied the interest of researchers in the areas of SLA, language assessment, and psycholinguistics. The underlying issue is that the human brain's cognitive processes (including linguistic ones) are not directly observable, and so it is not possible to

detect how L2 knowledge is represented, or at least what type of knowledge (implicit or explicit) a learner uses when they use an L2 language (although there are advances in neurolinguistic studies revealing a separation between the two types of knowledge, R. Ellis & Roever, 2018). Thus, the indicators of automatic processing required as an idiosyncrasy of implicit knowledge acquisition are inferable; that is, they can be observed indirectly via performance tasks. Ellis and Roever (2018, p. 1) added that "this requires making predictions about what kinds of behaviour are most likely to constitute evidence of a learner's implicit and explicit knowledge and then developing validity arguments to support the theoretical premises of these predictions".

A range of research studies on measuring the L2 implicit knowledge arising from factor analysis and correlational studies have attempted to establish the validity of a battery of tests hypothesized to provide comparatively independent measures of implicit and explicit grammatical knowledge (Bowles, 2011; R. Ellis, 2005; J.-e. Kim & Nam, 2017; Loewen, 2009; Sarandi, 2015; Spada et al., 2015). These studies have obtained consistent evidence that the elicited imitation task (EIT), the oral narrative task, and the timed GJT are possible measures of implicit grammatical knowledge, while an untimed GJT and a metalinguistic knowledge test are possible measures of explicit grammatical knowledge.

2.5.1 Elicited imitation task (EIT)

The elicited imitation task (EIT) was first used as a language-testing technique for assessing children with language delays or disorders (Erlam, 2009). In this test, utterances are read aloud to a subject, who then has to repeat them orally as accurately as possible. The utterances usually had targeted linguistic structures, and the subject's accuracy was a reflection of their knowledge of these structures and linguistic competence (Chaudron, 2003; Vinther, 2002).

This measure then found use in a wider range of research studies on L1 and L2 acquisition. Early results accepted the EIT as a measure of grammatical proficiency and language development, but other studies rejected the EIT; by arguing that EIT "is a perceptual-motor skill not dependent on comprehension" (e.g., Fraser et al. 1963, p. 483), or by affirming that the EIT can function as a measure of language production in experimental settings but not to measure spontaneous production (Prutting, Gallagher, & Mulac, 1975, see Vinther, 2002, for a comprehensive review). Vinther (2002) described the EIT as 'slippery' due to the ambiguity of the underlying neurolinguistic processes when a subject reproduces a given stimulus. The major controversy centred on subject processes and interpreting stimuli before repetition or simply using rote repetition to imitate without comprehension.

In the past decade, the EIT has gone through several redesigns in an attempt to construct a more reliable measure for implicit grammatical knowledge (e.g., Bowles, 2011; R. Ellis, 2005; Erlam, 2006; J.-e. Kim & Nam, 2017; Sarandi, 2015; Spada et al., 2015). This striking revival of interest in validating the EIT began after Munnich, Flynn, and Martohardjono (1994) identified two aspects to consider when redesigning it to maximize its validation (Sarandi, 2015). The first recommendation was that the subjects' attention should be directed at the meaning of an utterance before they are asked to repeat it. The underlying assumption here is that having a subject respond to the meaning under time pressure reduces the tendency for them to focus explicitly on the linguistic form and, hence, explicit linguistic knowledge is accessed. To manage this, a subject is asked to agree, disagree, or express no opinion about the content of an utterance (meaning processing), and then he/she is required to repeat the utterance. This means that the task requires both input and output processing – auditory comprehension and oral production.

The second recommendation was that the EIT include both grammatical and ungrammatical utterances rather than being limited to non-deviant utterances that the subject

should repeat in correct English. The underlying hypothesis is that the spontaneous and fluent correction of deviant structures is a strong enough indication that a subject has internalized these grammatical forms (Erlam, 2006; Munnich et al., 1994). This internalization reflects implicit automatic knowledge of these structures. R. Ellis and Roever (2018, p. 8) clarified that "learners draw on their long-term memory of the target structures and thus automatically correct the errors without necessarily noticing them."

The construct validity of measures of implicit and explicit knowledge was first sparked by R. Ellis (2005). He conducted a psychometric study to investigate the correlations between scores on 17 target structures in 5 language assessment tests: an EIT, an oral narrative, a timed GJT, an untimed GJT, and a metalinguistic knowledge test (see Section 2.5.2 for more information about the timed GJT). These tests were completed by 20 native speakers of English and 91 L2 learners of English. The EIT included the two recommended suggestions mentioned above: attention directed to the semantic meaning of the utterance (i.e., message processing) and including ungrammatical items. More specifically, 32 utterances were created which involved both grammatical and ungrammatical sentences where each sentence included a target structure. The stimuli were presented orally to the participants. To encourage them to process the utterances for meaning, they were first asked to judge whether they agreed with the content of each sentence. Next, the participants were required to reproduce the sentences orally in correct English. The responses were scored as '1' if the obligatory occasion using the target structure was correctly produced, and the response was scored as '0' if the target structure was either incorrectly produced or avoided. Average scores were expressed as percentage correct. The factor analysis results demonstrated that an EIT loaded heavily on a single factor together with an oral narrative task and a timed GJT. This indicates that the three measures have one similar trait. The author also concluded that, of the three measures of implicit knowledge, the EIT was the most

accurate measure of assessing implicit knowledge. This is because it is a demanding task which involves a mixture of listening comprehension and oral production; hence, this task calls for more language competence than oral narrative tasks (Sarandi, 2015). Consequently, the EIT has recently been subjected to empirical validation in an attempt to enhance its effectiveness in measuring implicit knowledge.

Ellis's (2005) results have been replicated with different learner populations (e.g., Bowles (e.g., Bowles, 2011; Zhang, 2015). Zhang (2015) conducted a validation study with university-level learners of English in a Chinese university to examine the extent of their implicit and explicit L2 knowledge. One hundred Chinese EFL students completed four tests: an EIT, a timed GJT, an unGJT GJT, and a metalinguistic knowledge test. The results of a factor analysis were consistent with Ellis's (2005) findings in that they demonstrated that the four measures tapped into two different types of knowledge (i.e., implicit and explicit knowledge). Zhang (2014) found that the EIT and timed GJT had more potential to elicit implicit knowledge and that the untimed GJT in ungrammatical items and the metalinguistic knowledge test were more likely to elicit explicit knowledge.

Erlam (2006) replicated Ellis's (2005) study but focused only on examining the validity of the EIT. The task included 17 grammatical structures, and each targeted structure had 2 utterances. The length of statements varied from 6 to 18 syllables and averaged 12.27 syllables. The study recruited 20 native speakers of English and 95 L2 English learners. The scores achieved in the EIT were compared with two timed language tests (oral narrative task and IELTS) and with the scores obtained from native speakers of English. The results indicated a significant correlation between the EIT and the listening and speaking sections of IELTS ($r = 0.67$, $r = 0.72$ respectively), and between the EIT and the oral narrative task ($r = 0.48$). The author succeeded in finding evidence that the EIT is a likely measure of implicit grammatical knowledge. The findings also revealed no significant correlation ($r = -0.28$)

between the length of EIT statements and the success of correct repetitions made by the participants. However, Saranda (2015) claimed that, although Erlam's (2006) study had a degree of validity, the data gathered by the study should be treated with caution. For example, the correlation between the EIT and the oral narrative task ($r = 0.48$) was computed according to 7 targeted structures in the oral narrative task and 17 targeted structures in the EIT test. When the correlations between the four structures used in both measures were investigated individually, only two structures revealed statistical significance, with a small correlation (third person '-s', $r = 0.42$, and regular past '-ed', $r = 0.36$). For the third structure, 'verb complementation', the statistical significance was slightly above chance ($p \leq 0.06$), and the correlation was also found to only a small degree ($r = 0.26$), which is considered a weak correlation. The correlation for the fourth structure did not reach significance. Moreover, Erlam (2006) applied only a very limited number of test utterances (two sentences) to each targeted structure in the EIT, resulting in the potential deficiency in the validity of correlation between the tests.

Despite the resounding success of the theoretical and practical work carried out by Zhang (2015), R. Ellis (2005) and Erlam (2006), which was based on a similar EIT design, these studies had some general salient limitations at the methodological level. First, the way the participants processed the utterance stimulus for meaning rather than imitating it by rote could still allow participants to use their explicit linguistic knowledge. Although researchers agree that oral tasks such as EIT more often reflect a higher use of implicit knowledge as "the signal is fleeting and not available for prolonged reflection" (Rebuschat, 2013, p. 612), the latency between the stimulus and its repetition, used for giving opinions about the content of the statement, could leave sufficient time for a subject to plan or think about the appropriate syntax, thus accessing explicit knowledge. R. Ellis (2005), therefore, strongly believed that the EIT should be administered under time constraints; this would reduce the possibility of a

subject planning or monitoring responses. The previous validation studies have suggested that the EIT is time pressured, where the utterance is presented to a subject in an auditory manner only once and in real time, and the subject is asked to indicate whether they agree with the content of the sentence and, finally, to repeat the utterance at their own pace. However, this procedure clearly indicates that there was no fixed time limit or control over the latency between prompt and response, and pressure time was left to the discretion of the administrator running the task.

The second methodological issue refers to the limited number of test items presented in the EIT. In total, 34 sentences were employed to target 17 grammatical features. Due to the inclusion of numerous features in the EIT, each one was tested on only two trials in the task, which may not be sufficient. Bley-Vroman and Chaudron (1994) suggested that "many samples of the target structure(s) are necessary in order to avoid the 'fixed effect fallacy' (cf.Clark, 1973, p. 258) of eliciting too few instances of the structure, in which unknown factors such as word familiarity or even phonological features might facilitate or inhibit imitation".

The two limitations - the number of test items and the latency between prompt and response - were addressed by Spada et al.'s (2015) validation study. The study employed one syntactic structure - 'passive structures' – and imposed a time limit for oral responses. A total of 93 participants, including 73 L2 learners and 20 native speakers of English, completed the EIT. Spada and her associates (2015) claimed passive structures were selected for their known difficulty in L2 acquisition. To establish the time limit in the EIT, participants only had 6 seconds after a stimulus to give their opinion. This was to prevent them resorting to rote memory and monitoring the rule governing the stimulus. They were then asked to repeat the sentence they had heard in correct English as soon as possible after they had heard it and before 8 seconds had elapsed. The results from the factor analysis showed that learners' EIT

scores significantly loaded on the factor labelled implicit L2 knowledge, hence validating the results of previous studies, which found that the EIT is a valid measure of implicit grammatical knowledge. However, it needs to be noted that diminishing the number of stimulus grammatical features to one stimulus could make it difficult to generalise the finding.

The aforementioned studies found that EIT is a reliable measure for implicit knowledge. However, Hulstijn (2002) theoretically argued that implicit knowledge cannot be as qualitatively described as automatized knowledge, and the two, as a result, should be distinguished from each other. Accordingly, Suzuki and DeKeyser (2015) supported this notion by empirically examining the validity of using the EIT and word monitoring tasks to measure implicit knowledge. Japanese L2 speakers with Chinese as their L1 performed the EIT task with the integrated word-monitoring component, a metalinguistic knowledge test, and a probabilistic serial reaction time task. The metalinguistic test assesses explicit knowledge, and the probabilistic serial reaction time task measures an aptitude for implicit learning—both of which served as a performance baseline for the EI task with the built-in word-monitoring component. The word-monitoring component incorporated into the EIT required participants to read a target word, and they were instructed to press a keyboard button as soon as they heard the target word in the auditory sentence. This task's underlying rationale is that response time slows down when a target word appears after an ungrammatical structure when compared to a grammatically correct structure. The difference between reaction times in the grammatical and ungrammatical items can indicate how much participants can detect the deviant structure in the utterances. Hypothetically, the magnitude of the index may possibly indicate a better measure of implicit knowledge. The results revealed a positive correlation between the EIT scores and performances on the metalinguistic knowledge task, but a lack of correlation was found between the EIT scores and performances on the probabilistic serial reaction time task. The word-monitoring performance, in contrast, was unrelated to metalinguistic knowledge but correlated positively with the

probabilistic serial reaction time task scores only among L2 speakers with longer residence times. The study claimed that online error detection can measure implicit knowledge (i.e., linguistic knowledge is deployed without awareness), whereas EIT may measure automatized explicit knowledge (i.e., linguistic knowledge is deployed with awareness even if algorithmic rules are executed rapidly or are highly automatized). Despite the significant results, this study's findings must be seen in light of three limitations, specifically regarding the conditions through which the EIT was designed and the lack of a subjective measure in the data collection. The first limitation concerns the instructions given to participants; they were instructed to comprehend utterances so they could answer the belief statement question, but they were asked to 'correct any ungrammatical utterances if necessary'. This instruction may stimulate the participant to focus on form, which thus increases the possibility of accessing the explicit knowledge of the target structure. The second limitation concerns the time limit for repeating the utterance—participants were allowed to repeat each sentence within 8 seconds regardless of the length of sentences. This might allow participants to monitor the target structure when repeating, specifically with short sentences. These two factors, thereby, might be sufficient to generate a correlation between EIT scores and performance on a metalinguistic knowledge test. The final concern relates to subjective measures where the study's conclusions cannot prove beyond a shadow of a doubt whether linguistic knowledge was accessed with or without awareness while performing the online error detection without using subjective measures.

To reiterate, previous studies have shown the resounding success of the theoretical and practical work (e.g., Bowles, 2011; R. Ellis, 2005; Erlam, 2006) in developing an EIT as a measure of implicit knowledge. Nonetheless, despite attempts by these validation studies to focus participants' attention on the meaning of utterances, and despite the inclusion of ungrammatical items, the time gap between the auditory stimulus and the oral imitation was not controlled, which could result in greater access to explicit knowledge. Spada et al. (2015) conducted a successful attempt to narrow the time lapse between the stimulus and the

subject's response. They allowed only 6 seconds for participants to give their opinion on the utterance and 6 seconds to give an oral response. The results demonstrated that the EIT tapped into implicit L2 knowledge. This time limit put more processing pressure on the subject, resulting in narrowing down the access to explicit knowledge. Finally, Suzuki and DeKeyser (2015) conducted promising work for distinguishing implicit from automatized explicit knowledge. However, due to the limitations in the study, definitive conclusions cannot be drawn.

2.5.2   Grammatical judgement task (timed GJT)

The GJT has been commonly employed in SLA-based research studies (e.g., Montrul, 2005; Nabei & Swain, 2002; Toth, 2006) and in standardized tests (such as the Test of English as a Foreign Language [TOEFL] paper version) to assess L2 learners' linguistic knowledge. However, concerns have been raised over the type of linguistic knowledge that the GJT measures: implicit, explicit, or some combination of the two (see Loewen, 2009, for a comprehensive review).

The GJT can be defined as a task in which subjects are required to "decide whether a sentence is well-formed or deviant" (R. Ellis, 1991, p. 162). The GJT was first used by linguists from the generative grammar tradition, such as White (1988) and Schachter (1989), to provide intuition about native-language sentences. The impulse for this trend is mostly a result of the application of Chomskyan grammatical theory (i.e., Extended Standard Theory and the Principles and Parameters model, see Davies & Kaplan, 1998 for a review). That is, L2 learners' judgements in the GJT were compared to native speakers' judgements where the latter were used as a baseline for comparison against which the linguistic competence of learners' judgements can be measured (Loewen, 2009). Furthermore, the GJT has been increasingly popular in SLA research studies, as such a task was assumed to examine which

grammatical structures are complex in output production as a result of their infrequent occurrences (e.g., R. Ellis, 2009b; Mackey & Gass, 2005).

Previous research has emphasized that the most influential design attribute of the GJT in triggering learners to access implicit knowledge is time pressure. The underlying assumption is that a rapid linguistic behavioural response to the GJT sentences reflects intensity of availability and automaticity of the acquired knowledge (Loewen, 2009). Put simply, "the faster the response time, the more "automatic" the decision" (Loewen, 2009, p. 98, as cited in Alanen, 1997).

In Ellis's (2005) study, the timed GJT consisted of 68 sentences containing 17 grammatical structures, with half of the stimuli being grammatical ($n = 34$) and the other half being ungrammatical ($n = 34$) where each structure received 4 sentences. The stimuli were presented visually on a computer screen. The time limit for judging whether the stimulus was grammatically acceptable or not was established based on native speakers' average response time for each sentence. The time limit for each sentence was increased by 20%. Ellis claimed that this additional time allows for the slower linguistic processing speed of L2 learners. Each stimulus was scored dichotomously as correct/incorrect, with items left with no response being scored as incorrect. The results of the study indicated that the timed GJT loaded together with EIT and an oral narrative task on one factor. This means that the time pressure inherent in the GJT triggered learners to access implicit knowledge. That is, the scores obtained from the timed GJT loaded on a factor which was different from the factor which the scores of untimed GJT loaded on.

Loewen (2009) conducted a research study to examine the construct validity of timed and untimed GJTs by testing them against the criteria for measuring explicit and implicit grammatical knowledge respectively. The timed and untimed GJTs consisted of an identical set of 68 sentences targeting 17 structures; each targeted structure was presented 4 times (in 2

grammatical and 2 ungrammatical sentences). In total, 140 L2 subjects were recruited, and 18

native speakers of English took part to form a control condition. The length of the sentences

was not controlled. The time limit for judging each sentence was established based on native

speakers' performance (see R. Ellis, 2005 above for more information about the time limit).

The results indicated that the timed GJT appears to highly reduce the likelihood that learners

access their explicit grammatical knowledge when making judgements. Further results also

indicated that learners performed significantly better on the grammatical than ungrammatical

items in both types of GJT.

Loewen's results supported the claim that responses to grammatical and

ungrammatical sentences draw on two distinct cognitive processes or databases that involve

accepting well-formed sentences and rejecting ungrammatical sentences. Ellis (1991, p. 178)

observed that "sentences that learners judged to be ungrammatical or that they were not sure

about often invoked attempts to make use of declarative knowledge".

Loewen (2009) in his study argued that if the hypothesis established by Ellis (2004)

was taken into account, that is, that a participant engages in a three-stage process when

performing the GJT: i) semantic processing, ii) noticing, and iii) reflecting, then it is likely

that grammatical and ungrammatical items are processed drawing on two different

knowledge sources. That is, the sentence first needs to be processed semantically in order for

its meaning to be unpacked (stage 1). Next, the sentence is examined syntactically for

deviation (stage 2). If no syntactic deviation is detected in the sentence, the participant can

make their grammaticality judgement at this point. If a syntactically deviant element is

detected, whether based on intuition or by explicit knowledge, or syntactic ambiguity is

present in the sentence, covering up the semantic meaning of the sentence, reprocessing and

further examination will be involved prior to the judgement (stage 3). On the other hand, if

the deviant element is obvious to a participant, he/she may immediately judge the sentence, and no further process is needed.

To add to the understanding of drawing on distinct sources of knowledge based on the type of grammaticality, Gutiérrez (2013) conducted a factor-analysis study to investigate whether responses to grammatical and ungrammatical sentences in both timed and untimed GJTs draw on different types of grammatical knowledge. In this study, 49 non-native speakers of Spanish completed the timed GJT. The results show that the grammatical and ungrammatical items of a timed GJT loaded differently in both exploratory and confirmatory factor analyses. The author proposed that grammatical items induce L2 learners to resort to their implicit knowledge, while ungrammatical items cause L2 learners to draw upon explicit knowledge.

However, J.-e. Kim and Nam (2017), in a study that presented the timed GJT in the auditory modality, discovered that responses to ungrammatical sentences loaded more strongly on an implicit knowledge factor than grammatical sentences. Ellis and Roever (2018, p. 11) concluded that "clearly, results for grammatical versus ungrammatical sentences have not been consistent across studies although the weight of the evidence supports Ellis' finding, namely that ungrammatical sentences are the more likely to elicit explicit knowledge".

2.5.3 Cognitive processing: timed GJT versus EIT

Responses to both the timed GJT and EIT reflect underlying automatic processing in a similar way (J.-e. Kim & Nam, 2017), where the most shared factor associated with the automaticity is speed of processing so that the possibility of accessing explicit knowledge is highly minimized. However, Segalowitz (2003) argued that although it is sensible to believe when a mechanism has become automatic, it functions faster than it did earlier, it does not necessarily suggest that it is thoroughly automatic. This is apparently manifested in the nature of distinct processing operations involved in the timed GJT and EIT. In the timed GJT,

attention must be focused on grammatical forms, and thus it is logically acceptable that the more rapid the response to the stimuli, the greater the degree of automaticity that was reflected. However, if the principle required in the timed GJT was applied similarly to the EIT, repeating the stimulus verbatim would most likely occur as a result. Therefore, in the EIT, attention must be focused on meaning, and there should further be a latency between the stimulus and its repetition in that this latency is used for giving opinions about the semantic content of the utterance. However, this technique should be conducted under time pressure so that the possibility of a learner having time to plan or think about the appropriate syntax is minimised.

In relation to the modality of each task, the GJT requires only visual processing (or only auditory processing if it is presented in an auditory mode) while the EIT requires auditory input and oral output processing (see Section 2.7 for more discussion about the difference between auditory and visual modalities). The oral production involved in EIT requires several processes that the L2 learners have to employ to accomplish a competent speech production: conceptualization, formulation, articulation, and self-monitoring (Bygate, 2001; Levelt, 1989). The speakers must first conceptualize the message content, relying on background knowledge and knowledge about the topic itself. Next, linguistic knowledge has to be formulated by discovering the relevant words and phrases and positioning them within the appropriate syntax. The third process involves ensuring speech motor control to articulate utterances such as movement of the lips, tongue, alveolar, palate, velum, and glottis. The last process is concerned with monitoring the output speech and self-rectifying mistakes. All these processes must be carried out automatically because humans do not possess sufficient attention capacity to control the four processes. Skehan (1998) suggested that there is potentially a trade-off between the accuracy, fluency, and complexity of speech. That is, allocating greater attention to one aspect (e.g., accuracy) might limit an L2 learner's capacity

for other aspects. The GJT, on the other hand, requires learners to access only their L2 orthographic lexicons (Fitzpatrick, 2007), where they may have a sufficient quantity of L2 lexical items to attain a reasonable level of reading competence, specifically when the target language is closely related to the first language (Ringbom, 2007).

Considering the different nature of the two tasks, the EIT is likely to impose greater processing demands on L2 learners' cognitive resources than the processing of visual text (i.e., GJT), resulting in poorer performance in the EIT. This theoretical interpretation is empirically supported by the previous studies, which found that learners achieved a greater performance on the GJT than the EIT (e.g., Bowles, 2011; R. Ellis, 2005; Spada et al., 2015; Zhang, 2015). Accordingly, it is possible to infer that the EIT and timed GJT possibly make different processing demands on accessing implicit knowledge.

J.-e. Kim and Nam (2017) conducted a factor-analysis study to examine whether the level of implicit grammatical knowledge tapped by an EIT is different from that tapped by a timed GJT. In total, 9 native speakers of English and 66 adult learners of English were recruited in the study. The time limit for the timed tests was established based on native speakers' performance. Each non-native speaker participant completed five tasks: a timed GJT presented in visual mode, a timed GJT presented in auditory mode, an EIT in which time was controlled, an EIT in which time was uncontrolled, and a metalinguistic knowledge task (MKT). The scores of the tests loaded largely on three different factors: the timed and untimed EITs on one factor, the auditory and visual timed GJTs on a second factor, and the MKT on a third factor. However, the EITs and the timed GJTs loaded heavily on the factors (Factors 1 and 2) that were not heavily loaded on by the MKT. This implies, as the study suggested, that the MKT taps on a different type of knowledge (i.e., explicit knowledge), which increases the construct validity of the EIT and the timed GJT as measures of implicit grammatical knowledge. The study explains that the distinct factor loadings of the EIT and

the timed GJT stem from the difference in the depth of processing between the two tasks, suggesting that the EIT has better access to implicit knowledge than the timed GJT. The results further revealed that the timed EIT and the timed GJT presented auditorily measured stronger implicit knowledge.

To summarise, the review of the literature pertaining to EIT and timed GJT indicates that the responses to both tasks (EIT and timed GJT) reflect underlying automatic processing in a similar way, which presumably are measures of L2 implicit grammatical knowledge. The EIT involves auditory input and oral output processing while the timed GJT requires only visual processing. Thus, the former requires a deeper level of processing than the timed GJT's shallower, perceptual level, resulting in the EIT imposing greater demands on cognitive processing resources than the timed GJT. Previous studies have indicated that the EIT is better at tapping into stronger implicit knowledge (i.e., reflected in automaticity) than the timed GJT. In the EIT, particularly, two factors should be involved to reduce the tendency to access explicit knowledge: first, learners' attention should be primarily allocated to meaning rather than form, and second, the time gap between the stimulus and its repetition, which also included the latency used for giving opinions about the content of the statement, should be controlled. In addition, the EIT should include both grammatical and ungrammatical utterances where the spontaneous correction of syntactic anomalies is a strong and sufficient indication that automatic language processing is in operation (i.e., utterances are not repeated by rote).

In the case of the GJT, in contrast, it requires a participant to judge under time constraints, from the first impression, whether a visually presented sentence is grammatically acceptable. Compared to the EIT, the most influential factor in the GJT in terms of inducing learners to access implicit knowledge is the amount of time participants are given to make

judgements. That is, a rapid linguistic behavioural response to the GJT sentences reflects intensity of automaticity of the acquired knowledge.

2.6   Applying implicit knowledge measures in L2 acquisition research

Following the development of the valid competency measures (specifically, the EIT and the timed GJT) to access implicit grammatical knowledge, SLA researchers have demonstrated a burgeoning interest for examining the learning mechanisms which might foster the acquisition of L2 implicit knowledge. These mechanisms included learning conditions such as explicit instruction (Akakura, 2012), enriched vs. enhanced input (Reinders & Ellis, 2009), input vs. output (Erlam, Loewen, & Philp, 2009), and implicit vs. explicit corrective feedback (R. Ellis, Shawn, & Erlam, 2009). Two types of learning conditions, namely, focus on form and oral output, will be addressed below; these are presumed to promote the acquisition of L2 implicit knowledge.

2.6.1   Focus on form

Focus on form is derived from the research investigating the role of attention in language acquisition (Pietrzykowska, 2011). The mechanisms involved in first language acquisition are not sufficient for L2 acquisition due largely to learnt attention and L1 transfer (N. Ellis, 2011). Thus, some degree of attention to linguistic form is necessary to speed up the acquisition of L2. Long (1991, pp. 45-46) defined focus on form, stating that it "overtly draws students' attention to linguistic elements as they arise incidentally, in lessons where the overriding focus is on meaning or communication". Long (1991) and Doughty (2001) have claimed firmly that focus on form is best employed to foster interlanguage development "because the acquisition of implicit knowledge occurs as a result of learners attending to linguistic form at the same time they are engaged with understanding and producing meaningful messages" (R. Ellis, 2006, p. 101). A central characteristic of 'focus on form'

learning is that it emphasises form–function mapping. That is, it spurs learners to "connect grammatical forms to the meanings they realise in communication" (R. Ellis, 2006, p. 101), and as a result, learners can acquire the L2 target structures (the desired objective) naturalistically while they process the input for meaning (Hulstijn, 2008 ; Reinders & Ellis, 2009 for review). However, it is crucial to maintain a distinction between focus on form and focus on forms. While the former requires the primary attention to focus on meaning, the latter involves primary attention to focus on linguistic forms (Long & Robinson, 1998).

One common technique to increase the likelihood of the target structure being noticed is by manipulating the input in order to attract learners' attention to the target feature, for instance, by underlining, capitalising, or putting it in bold (enhanced input) or by artificially increasing its frequency in the input (enriched input). Based on Schmidt's (1990, 1992) Noticing Hypothesis, the high frequency of sentences (i.e., enriched input) containing the target structure in the context of a meaning-focussed activity where primary attention is on semantic or pragmatic meaning (i.e., on message processing) is predicted to induce the learner to notice the target structure (i.e., the learners' attention is attracted by the frequent feature) and, thus, is successfully acquired and easily internalised. Noticing in the input enhancement refers to "the cognitive activity that learners engage in when they consciously attend to some linguistic feature in the input" (Reinders & Ellis, 2009, p. 283). Reinders and Ellis argued that when the target feature is noticed, it will be rehearsed in the short-term memory and therefore, this will improve the potentiality of acquiring it (i.e., integrating it into interlanguage). Loewen (2014) also suggested that attention to both meaning and form could create an optimal learning technique in which both implicit and explicit learning occurs and in which implicit and explicit knowledge are promoted.

A review of the previous research studies provides evidence to indicate that a focus on form results in the development of language learning (see Nassaji, 2013, for a comprehensive

review). However, research examining the impact of focus on form on L2 acquisition (i.e., implicit knowledge) is very limited (e.g., Pietrzykowska, 2011; Reinders and Ellis, 2009).

Pietrzykowska (2011), for instance, examined the relative effects of input enhancement on the L2 implicit and explicit knowledge acquisition of English embedded questions using a pretest-posttest design. The participants (n=56) were assigned to one of two groups: experimental (n= 32) and control (n=24) groups. The experimental group was exposed to three reading texts followed by comprehension questions in which the embedded question was made more noticeable in the texts through three typographical cues; that is, a different type of font, a different size of font, and boldfacing. The frequency of the embedded question was the same in each text where each text had eight items. The control group, on the other hand, was exposed to the same comprehension texts, but the target structure was not emphasised in any way. However, the number of embedded questions was decreased by half. The testing stage included two types of measures, specifically, a timed GJT and an oral narrative task, as possible measures of implicit knowledge, and an untimed GJT and a paraphrasing task, as possible measures of explicit knowledge. Although most of the results did not demonstrate statistical significance, the authors concluded it is the implicit knowledge rather than explicit knowledge that is more likely to be promoted by the type of input enhancement. Despite the interesting work performed in this study, caution should be used when interpreting these results. In the timed GJT, for example, the time limit was not established based on native speakers' average response time for each sentence; rather, it was based on the length of a sentence (i.e., it ranged from 3 to 5 seconds). In addition, exposure to 24 exemplars of the target structure might not be sufficient to effectively enhance the acquisition of embedded questions.

To better understand empirically the relative effects of enriched and enhanced input, Reinders and Ellis (2009) exposed participants to materials targeting negative adverbials with

subject–verb inversion (e.g., *never does my wife have friends over for a drink*) in two different learning conditions (enriched input vs. enhanced input), which both constitute 'focus on form'. In the enriched input, participants were exposed to "input where a specific L2 feature occurs with high frequency" (Reinders & Ellis, 2009, p. 282). It was hypothesized that exposure to high frequency would induce participants to notice the target structure and thus acquire it more easily. The enhanced input is rather similar to the enriched input; however, the target structure is emphasized in some way, such as glossing, boldfacing, or underlining. In both learning conditions, the participants were asked to complete three types of learning tasks: a dictation task, an individual text reconstruction task, and a collaborative reconstruction task. These tasks, generally, were in the form of texts about a range of general interest topics and had been seeded with a high frequency of the target structure. In the reconstruction task, for instance, participants were occupied in an activity which required listening twice to a passage of 60 - 70 words that was seeded with several instances of the target feature. Participants were not informed about what to look out for, but they were allowed to take notes while listening. Then, they were asked to reconstruct the heard passage by writing it out. In the enriched condition, participants performed the four tasks in a complete absence of grammar instruction or cues, while the enhanced input included some explicit hints about the syntax governing the negative adverbs. An example of the instruction provided to participants in the enhanced condition is as follows:

> *Listen carefully and pay attention to where the auxiliary verb comes in each sentence. For example, in the sentence 'Rarely has so much rain fallen in such a short time', the auxiliary is 'has', and it comes before the subject of the sentence 'so much rain'.*

The results revealed that participants in the enriched input condition outperformed those in the enhanced condition in relation to grammatical sentences only (the test included

grammatical and ungrammatical sentences) in the timed GJT. Conversely, no performance differences were found between the two conditions for the untimed GJT (the measure of explicit knowledge). However, the study had a salient limitation of the small sample size where the enhanced condition group included only 11 participants while the enriched group included 17 participants.

In summary, it is focus on form that is most likely to engender the acquisition of implicit knowledge (Doughty, 2001; R. Ellis, 2006; Long, 1991).This is premised on the assumption that linking grammatical forms to the meanings the learners comprehend from the input has facilitative effects on L2 acquisition. Previous empirical research on the development of L2 learning has provided evidence to indicate that focus on form plays a crucial role in the L2 learning development. Although studies (e.g., Pietrzykowska, 2011; Reinders & Ellis, 2009) that examined the effects of focus on form on the acquisition of the implicit knowledge of L2 are limited and further included some limitations, their results can indicate that focus on form is able to promote learners' grammatical competence (i.e., implicit knowledge).

2.6.2   Output production

Despite the broad theoretical recognition of the significant role of input in L2 acquisition (Krashen, 1989; Larsen-Freeman & Long, 1991; Schwartz, 1993; VanPatten, 2004), the role of oral output practice remains a controversial issue in L2 acquisition research in terms of whether it plays a facilitative role in the success of the L2 acquisitional process. Krashen (1998) argued that output does not play a significant role in the development of linguistic acquisition because i) output, specifically comprehensible output, cannot be readily available, ii) language acquisition might be successfully developed in the absence of output, and iii) there is no supportive evidence to indicate that output practice results in successful language acquisition (see Muranoi, 2007, for a comprehensive review). VanPatten (1996)

further argued that if the input is ignored (or downplayed) during learning while output practice is involved, the case is "analogous to attempting to manipulate the exhaust fumes (output) of a car to make run it better" (p. 6). Later, VanPatten (2004, p. 43) concluded:

> We are currently unable to support any specific role for output in the creation of an underlying competence that contains form-meaning connections. At best we can say that input is necessary for acquisition, but input and output may be better—we just don't know how or under what circumstances.

Similarly, R. Ellis (1994; 1997, cited in Muranoi, 2007) claimed that output practice plays only a limited role in L2 acquisition because there is not sufficient evidence that output practice contributes in any way to the success of L2 acquisition, although it might partially assist in increasing learners' consciousness of certain grammatical structures.

Contrary to these views, several output advocates claimed that output practice crucially contributes to L2 acquisition development. Swain (1985, 2005) asserted that output practice can play an ancillary role in L2 acquisition, specifically if cognitive processing is allowed (e.g., noticing, hypothesis testing, syntactic processing, and metalinguistic reflection). She further suggested that the role of output in L2 acquisition is not only limited to exercising existing linguistic knowledge but is also used as an approach to generating new linguistic knowledge. Toth (2006) suggested that it would appear an exaggeration to argue that input is the only force behind learning a new structure. Instead, a more compelling explanation is that the push to syntactically encode what the learner means to say, in conjunction with input from others and the learner's own conscious reflection regarding the acquisitional process of second language, together drive the learner's grammatical knowledge enhancement. Others, such as de Bot (1996) and Gass (1997, 2018), placed great significance on output by asserting that it can increase the degree of automated knowledge of L2 syntax. McLaughlin (1987, p. 134) argued that automatic processing involves "a learnt response that

has been built up through the consistent mapping of the same input to the same pattern of activation over many trials". Gass (2018, p. 148) extended this assumption to output, suggesting "the consistent and successful mapping (i.e., practice) of grammar to output results in automatic processing".

A large number of research studies (e.g., Benati, 2001; Cardierno, 1995; VanPatten & Sanz, 1995) sought to examine whether L2 acquisition is output-dependent. These studies, as Erlam et al. (2009) indicated, demonstrated that those who were exposed only to input performed better on interpretation tasks than those who were engaged only in output-based activities. They further revealed that these same participants, who had only input exposure, performed as well as those who had worked only at output-based learning on production tests.

Benati (2001), for example, compared the relative effects of two types of learning (comprehension-based input vs. production-based output) on the acquisition of future-tense verb morphology in Italian. The participants included 39 second-semester Italian students at the University of Greenwich. The comprehension-based input learning involved rule instruction and comprehension practice directed at shifting the way L2 learners process input to make better meaning–form connections. That is, rule instruction was presented first, then learners were engaged in aural and written activities (requiring matching and binary options) to respond to the content sentences. The output-based learning treatment included rule instruction followed by written and oral practice (part of which was meaning-oriented). Three language tests were developed for data collection, including an aural interpretation task, a written completion text, and an oral limited-response production task. The findings of the study indicated some evidence that input comprehension has positive effects on the acquisition of verbal morphology in Italian and better effects on the developing linguistic system of beginner L2 learners than production-based output learning.

However, the evidence supporting a greater role for output production in the acquisitional process of L2 language is derived from experimental studies that investigated output practice combined with input (e.g., Izumi, 2002; Morgan-Short & Bowden, 2006; Song & Suh, 2008; Swain & Lapkin, 1995). In Song and Suh's (2008) study, the role of output using two different types of output tasks (i.e., a reconstruction task and a picture-cued writing task) was investigated in terms of its efficiency in noticing and learning the past counterfactual conditional in English. A total of 52 intermediate Korean-English students were recruited to participate in the study. They were randomly assigned to two experimental groups (n = 29, n = 12) and a control group (n = 11). The treatment in the experimental groups included exposure to input and output practice. However, the control group received only input treatment involving reading comprehension. Overall, the results indicated that more noticing (i.e., the input is manipulated in a way to attract learners' attention to the target feature) occurred for students who had the opportunity to engage in output practice compared to those who did not. With respect to the promotion of acquisition, the results revealed that students who received output practice during the study performed significantly better than those in the non-output condition on the production posttest, but no difference was observed in the comparative effectiveness of the two output tasks. However, the study's limitations included a very small size, specifically in the experimental group that received a picture-cued writing task (n = 12), and the control group (n = 11). Such a small size minimised the power of the study and increased the margin of error.

In the same vein, a computer-based experimental study accomplished by Morgan-Short and Bowden (2006) examined the effects of meaningful (i.e., referential and affective activities) input- and output-based practice on L2 acquisition (a replication of VanPatten & Cadierno, 1993). The target grammatical feature was Spanish preverbal direct object pronouns. A total of 45 participants in a first-semester Spanish course at a university in the

United States were recruited in the study. The participants were randomly assigned to one of three groups (two experimental groups and one control group), which each contained an approximately equal number of participants. The two experimental groups (meaningful input- and output-based practice) were identical in terms of learning structure and content, but they differed in terms of the mode of the practice activities (aural for processing instruction and oral production for meaningful output-based instruction). The control group was exposed to the target structure through reading passages and comprehension questions. The participants' linguistic development was measured by assessing their interpretation and production of preverbal direct-object pronouns in Spanish. The findings indicated significant gains for the experimental groups regarding immediate and delayed interpretation and production tasks. It was also found that, overall, both experimental groups outperformed the control group on the interpretation task. For the production task, greater improvement was obtained by the meaningful output-based group than by the control group. Thus, the authors concluded that input-based as well as output-based learning can result in linguistic development. Nevertheless, the findings of this study must be considered in light of its limitation; the group that received output practice was not a pure output treatment group, as the participants in this group were exposed to input in the instruction and in the recast. As a consequence, the study could not deduce the effects of output that were separated from the recast.

To add to the understanding of the role of output in language acquisition, Erlam et al. (2009) investigated the relative roles of input- and output-based instruction in the acquisition of implicit language knowledge. The target structure was the use of indefinite articles for marking generic references. The subjects in the first of two groups sat for eight meaning-based output activities using a PPP (presentation/ practice/ production) format, which was adopted from Gower and Walters (1995). The second group sat for eight structured input activities. Both received the same explicit instructions. The testing stage included EIT as a

measure of implicit knowledge and untimed GJT as a measure of explicit knowledge. The results indicated that output-based instruction resulted in a significant improvement from pre- to posttests on the EIT. This indicates that learning conditions which give learners the opportunity for output experience result in the acquisition of implicit knowledge. However, the results indicated no significant differences in improvement between the two groups on the EIT. The authors attributed the lack of significant differences between the two groups to the possible lack of difference in the treatments each group was exposed to. However, they confirmed that the significant amount of learning (from pre- to posttests) resulting from the output is the most interesting feature, as it challenges claims that output does not have a specific role in building up an underlying linguistic competence (VanPatten, 2004). Nevertheless, it is worth noting that the participants' implicit knowledge was assessed using only the EIT; the timed GJT was not involved. Similarly, the intervention employed in this study was accomplished through classroom-based teaching that might not allow each subject intensive exposure to target structures and might also include threat factors, which could disrupt the process of learning.

To summarise, although the role of output practice was believed to be a contentious issue in L2 acquisition research, as shown by several ongoing debates, experimental studies (e.g., Morgan-Short & Bowden, 2006; Song & Suh, 2008) in which the L2 learners were engaged in output practice in conjunction with input rehearsal provided supportive evidence for the role of output production in the acquisition of L2. More specifically, Erlam et al. (2009) and Rebuschat & Williams (2012) convincingly contributed to the increasing evidence suggesting that output can promote a linguistic development system. In the former, the L2 learners who were exposed solely to output-meaning practice and received no input-meaning practice made significant gains from the pretest to both the posttest and delayed posttest on the EIT. In the latter, on the other hand, the experiment that employed oral repetition as an

additional training task produced a greater learning effect in terms of unconscious implicit knowledge than the experiment in which elicited imitations were absent. Therefore, the present study provided participants in auditory and visual groups with an opportunity to engage in language output in the form of elicited imitations as an additional training task in combination with input practice (see Section 3.4 for more information about the research design).

2.7   Input modality and language acquisition

The significance of sufficient input in driving the acquisitional process of a second language is indisputable. However, the success of L2 acquisition might be affected by the modality in which input is presented. Psychological research has provided evidence to suggest that input processing is significantly affected by the modality of input presentation (e.g., Broadbent, Vines, & Broadbent, 1978; Penney, 1989; Rollins & Hendricks, 1980; Shaffer, 1975). Penney (1989), for instance, had participants exposed to stimuli presented in two modalities, two languages, and based on distinct semantic categories. Penney found that participants preferred to recall the stimuli on the basis of the modality of input before the language of presentation or the semantic category. Thus, Penney elucidated the nature of the separate processing streams, hypothesizing that auditory and visual inputs are represented and processed in two separate streams, with each stream possessing its own properties and capabilities.

It has been hypothesized that visual stimuli must undergo articulatory rehearsal in order to be encoded in a phonological format and moved into the phonological store. In contrast, it is not necessary for auditory stimuli to undergo such re-encoding (transformation), as they are directly encoded in a phonological format (see Shallice & Vallar, 1990, for a review). That is, auditory material gains a direct and automatic access to the phonological working memory (more detailed information about working memory is provided under

Section 2.8). This is unlike the visual input, which involves three sequential processes, namely, i) *recoding*, grapheme-to-phoneme transformation, ii) *rehearsal*, which conveys visual stimuli to a storage system following the phonological recoding; and iii) *storage,* where the phonologically recoded material is provisionally maintained and where it is exposed to loss by deterioration or interference (Papagno, Lucchelli, & Vallar, 2008). The rehearsal process involved in processing visual stimuli revives the memory trace, and, as a result, the decay of the trace is inhibited (Baddeley, 1990). Therefore, the difference in the encoding modality presumes that the recall of verbal information is better when such material is presented visually rather than auditorily.

A crucial superficial aspect contributing to distinction in language performance depending on the encoding modality is referred to as input access. Studies (e.g., Alderson, 2005a; Anderson, 1980; Danks, 1990; Rost, 1990) found that access to language auditory input is more challenging than access to visual input. The common explanation refers to the distinctive idiosyncrasies of each modality. Anderson (1980) affirmed that the visual input streams are more easily systematically segmented into units than the auditory input streams. This is because the visual input is presented in the form of text that does not vary according to the author. That is, the orthographic representations of lexical items are immutable, and therefore, the visual input stream is segmented in the way that many written language systems are (Alderson, 2005a). Listeners, on the other hand, must deal with prosodic and intonational cues that vary from speaker to speaker to find word boundaries within a sentence. As a result, a heavy demand is placed on the processing resources, which might lead to unsuccessful auditory comprehension.

Another significant factor in the distinction between modalities in input access is that visually presented language allows for recursive processing and, thus, the rate of input can be controlled. This offers the possibility of concentrating and disregarding specific items of

input. Conversely, language input that is auditorily presented does not allow for this control and re-examination since auditory input is transient and must be processed in real time. That is, listeners must segment a continuous aural stream while simultaneously reacting to new input material (see Rost, 1990, for a review).

According to Perfetti's verbal efficiency theory (1985, as cited in Wu, 2016), word decoding (input access indicated as low-level process) and comprehension (high-level process) consume attentional resources. More specifically, the more attentional the cognitive resources imposed by the word decoding process are, the fewer residual resources are available for comprehension processing. L2 learners with highly efficient word recognition do not have to allocate more attentional resources to word decoding, and, consequently, more resources remain available for the processing of comprehension. On the contrary, those with slow word-recognition skills need to allocate their attentional resources between word decoding and comprehension, and consequently, comprehension is detrimentally affected. According to this theory, competence at word-level processing, is crucial to the processing success of visual linguistic input. This is because it is the efficiency of the input access processing which determines the residual resources available for learning and comprehension processing.

Despite the distinctive inherent idiosyncrasies of each modality, both involve achieving a complementary relationship between top-down and bottom-up processing in constructing comprehension. According to Richards (2015), Treiman (2003), and Vandergrift (2007), the top-down process requires an individual to have recourse to background and contextual knowledge, including knowledge of the topic and knowledge of the specific textual or aural contents, to construct a conceptual framework for comprehension. A schema is activated by the reader/listener to establish what prior knowledge they have. To this end, elicitation of information is "guided by individual's prior knowledge and expectations"

(Treiman, 2003, p. 664). The bottom-up process, on the other hand, is concerned with the decoding and processing of the auditory or visual input into basic language units (linguistic knowledge). The distinction between the two modalities lies in the bottom-up processing. In the case of visual input, the L2 reader must decode and encode lexical items, and combine words into phrases then sentences, which calls for semantic, syntactic, and pragmatic processing. By contrast, in the case of auditory input processing, such input involves "perceiving and parsing the speech stream at increasingly larger levels beginning with auditory-phonetic phonemic, syllabic, lexical, syntactic, semantic, propositional pragmatic and interpretive" (Field, 2003, p. 326). Field (2003) provided some examples of problems that an L2 learner might have when processing auditory input; for instance, an L2 listener might not distinguish between *want* and *won't* (a phonemic problem), or might fail to observe/ or disregard the semantic implications of the phoneme /v/ in a sentence such as *I've lived* (a syntactic problem). In Detey's words (2005, as cited in Detey, 2009, p.180), "not only do the linguistic properties of oral and written stimuli differ, but so does their psycholinguistic processing by language learners, from low-level perceptual process (reading process vs listening process) to higher-level mnemonic encoding in the phonological and/or orthographic lexicon." Overall, the bottom-up process 'linguistic knowledge' and the top-down process 'background knowledge' are mutually interactive and complementary in both modalities.

Research on the significance of input modality on L1 processing dates back to the 1960s (e.g., Dennis, 1977; Shaffer, 1975; Treisman, 1969). This research has indicated that adults are inclined to comprehend visual language input more effectively than auditory input. Daniel and Woody (2010), for instance, randomly assigned 48 university students to either read an article or listen to the same material in order to examine the retention of those who were auditorily exposed to the material and those who were visually exposed to the source as

text. The results showed that participants who received visual exposure scored significantly higher on a quiz than those who were auditorily exposed to the identical article. However, the study had a potential limitation. The treatment stage in the study required participants to read or listen to the material in uncontrolled conditions: They were instructed to read or listen to the material as part of their day for two days before the quiz. This means that each participant read or listened to the article in different circumstances, and this increased the threat to the internal validity of the study.

A recent study conducted by Rogowsky, Calhoun, and Tallal (2016) empirically examined the impact of input modality on comprehension in native English-speaking college graduates. In total, 91 participants were randomly assigned to one of three groups (i.e., auditory, visual, and dual modality) and given identical input materials. An immediate comprehension test was administered in written form, which included 48 multiple-choice questions designed to measure the recollection of the 2 passages. The same test was run 2 weeks later to assess their retention of the information. Participants in the visual modality read the material at their own pace, without time pressure. Re-exposure to the auditory or visual materials was prohibited. The results demonstrated no significant differences regarding the effects of the three different modalities on comprehension in either the immediate or follow-up test. However, a limitation of this study suggests caution in interpreting its results. The non-fiction material used in all groups was narrative in style and may not have represented the discourse style typically found in textbooks. Therefore, it could be that a similar experiment using textbooks, with the goal of learning or obtaining specific information, may generate different findings.

In L2 research, a number of studies have investigated how modality plays a differential effect on L2 performance. Johnson (1992) and Murphy (1997) examined the effects of modality on learner performance using an untimed grammatical judgement task.

Murphy's (1997) study sought to investigate the accessibility of universal grammar in SLA,
hypothesizing that some non-linguistic elements could affect subjects' performance on
metalinguistic tasks. The study investigated whether participants would judge auditory
statements differently from visual ones. The results indicated that subjects in the auditory
modality were slower and less accurate on the GJT than subjects in the visual modality. The
author concluded that L2 learners can be affected not only by the syntax of the target
sentences but also by the modality of the stimulus presentation.

To add to the understanding of the modality effects on L2 performance, Wong (2001)
investigated the impact of modality (visual vs. auditory) on input processing based on the
allocation of attention (meaning vs. form). The study had 79 French learners of English as a
foreign language take part in the experiment. The results supported VanPatten's (1990)
conclusion that attentional capacity is distinctly constrained in auditory versus visual
processing. More specifically, attention to the linguistic form of the input competes for
learners' limited attentional resources with attention to meaning in the auditory modality,
specifically during the early stages of L2 acquisition, resulting in a detrimental effect on the
comprehension of the input. However, attending to form and meaning simultaneously in the
visual mode does not negatively affect comprehension. The author suggested that this parallel
focus might be because the visual mode does not require more cognitive effort compared to
the auditory mode. The author also concluded that input modality is a variable that plays a
crucial role in performance when linguistic input is processed.

However, Morgan-Short et al. (2018) argued that VanPatten's (1990) and Wong's
(2001) research had some methodological limitations which could have affected the internal
validity of their studies. One of these limitations is linked to the tasks set (paper-based tasks)
in the visual modality stream. Morgan-Short and her associates claimed that when the
participants were instructed to answer comprehension questions after they had been visually

exposed to the Spanish passage for comprehension, they were not instructed that they did not refer back to the sentences, paragraphs, or the full passage. This means that the participants had the opportunity to re-inspect the passage to note grammatical forms, while such an opportunity was completely absent in the auditory tasks. Therefore, the authors conducted a multi-site replication study posing an identical research question, but this time, they controlled backtracking and included a time limit for the visual tasks in order to increase the comparability between the written and aural modalities. The study focused on Spanish learners' acquisition of English as a foreign language where participants were recruited from different countries. The results indicated attending to meaning and form simultaneously had no effects on comprehension in either modality. However, they suggested that further investigation be carried out due to site-specific effects and methodological limitations.

Park (2004) further examined the difference between L2 auditory comprehension and L2 visual comprehension with a focus on the roles of linguistic knowledge, background knowledge, and question types: global questions (i.e., listening or reading for main ideas) and local questions (i.e., listening or reading for specific details). A total of 168 Korean learners of English completed four assessments: a listening comprehension test, a reading comprehension test, a linguistic knowledge test (consisting of grammar and vocabulary), and a background knowledge test. The listening and reading tests had the same content, but the differences lay in the pattern of modality. The background knowledge test was employed to control the participants' background knowledge about the four passages used in the listening and the reading comprehension tests. The analysis of data showed a significantly better performance on the reading comprehension than on the listening comprehension test. In particular, participants processed inferential information more effortlessly than factual information in the listening comprehension test, while the opposite result was obtained for L2 readers. The results further indicated that linguistic and background knowledge produced

significant effects on the listening comprehension. In L2 reading comprehension, linguistic knowledge exerted a significant effect, while background knowledge played only a moderate role. The fact that linguistic and background knowledge as well as global and local strategies were used differently by participants in L2 listening and L2 reading comprehension led the authors to conclude that the two modalities are processed differently.

Acknowledging that modality has a differential effect on L2 performance, a limited number of studies have addressed the different effects of auditory versus visual input on L2 learning itself (e.g., Sagarra & Abbuhl, 2013; Sydorenko, 2010); to date, the results have been inconclusive, and some have revealed counter evidence (i.e., an advantageous bias towards auditory input). Sydorenko (2010), for instance, examined the effect of input modality on the vocabulary learning. The participants in this study were 26 learners of Russian, and they were assigned to one of three groups: a group exposed to video with audio and text (i.e., dual modality), a group exposed to video in the absence of caption texts, or a group exposed to video with only text (i.e., the video lacked audio). The caption texts on the screen were a transcription of the audio texts (i.e., the same language). Then, written and aural vocabulary tests and a final questionnaire were administered to participants. The findings indicated that groups with dual modality and text-only captions performed better on visual than on auditory recognition tests of vocabulary, while the groups with audio-only scored higher on auditory than visual tests. The questionnaire responses revealed that most of participants' attention was drawn to the text, followed by video and audio. The results of Sydorenko's (2010) study concur with those of other studies including those by Chai and Erlam (2008) and Jensema, Danturthi, and Burch (2001). In the former study, the results revealed that when captions were included alongside a (narrated) video, the learners' attention was inclined to be allocated to the visual text more than to the auditory input, which could hinder the processing of auditory material. However, although these studies support the

hypothesis that the inclusion of visual texts in video enhances language learning, as learners process visuals to confirm the precision of what they have heard, the processing of auditory and visual language inputs simultaneously could induce cognitive overload and seriously impair comprehension and learning (Mayer & Moreno, 1998; PujolĂ, 2002, see Wenhua, 2017, for a review).

Sagarra and Abbuhl (2013) compared the effects of auditory versus visual feedback on L2 learning in response to noun-adjective gender or number agreement errors. A total of 218 learners of Spanish were exposed to one of six types of feedback: visual utterance rejection (i.e., participants read *Wrong. Move on to the next sentence*), auditory utterance rejection (i.e., they listened to *Wrong* and read *Move on to the next sentence*), visual recast (i.e., they read *La silla es blanca. Move on to the next sentence*), auditory recast (i.e., they listened to *La silla es blanca, one three* and read *Move on to the next sentence*), visually enhanced recast (read *La silla es blancA. Move on to the next sentence*), and auditorily enhanced recast (listened to *La silla es blancá, one three*, with stress on the adjectival ending and read *Move on to the next sentence*). Following the treatment stage, participants first completed three written tests and then two oral face-to-face interaction tests. The results revealed that auditory enhanced recasts were more beneficial than typographically enhanced recasts on the interactional posttests only (but not on the written posttests). The authors claimed that although processing transient auditory input (as on the interactional posttests) is more cognitively demanding than visual input, "learners may have fewer available cognitive resources and thus may benefit from hearing prosodically highlighted recasts that clearly demarcate the nature and the location of the error while simultaneously providing a model of the target structure" (Sagarra & Abbuhl, 2013, p. 210). However, it should be pointed out that the study suffered from a potential internal threat that pertained to the testing order, leading to

greater accuracy on the written tests. That is, exposure to three written tests in the beginning might prime participants and, as a result, affect their responses in the oral tests.

However, empirical studies in cognitive psychology examining the effect of input modality on the immediate recall of verbal information (e.g., a list of digits, characters, or pseudo-words) had contradictory results and a different explanation (e.g., Avons & Phillips, 1980; Rummer & Engelkamp, 2001; 2003, see Rummer, Schweppe, & Martin, 2013, for a review). They revealed evidence that the immediate recall of verbal information is better when such material is presented auditorily than visually. This effect has been explained by the two types of storage existing in the sensory memory: the Pre-categorical Acoustic Storage, advanced by Crowder and Morton (1969) and the Pre-categorical Visual Storage, advanced by Sperling (1960). Generally, the sensory memory refers to "the registers for veridical, modality-specific representations that provide the first memories of inputs to each of our senses" (Lawson, Fernandes, Albuquerque, & Lacey, 2015, p. 231). Such a type of memory retains the first impressions of a stimulus (acoustic or visual) after it has ceased. It lasts only milliseconds and occurs mostly unbidden and without any conscious attention (Nursey & Phelps, 2016). Sensory memory cannot be prolonged through rehearsal, and as a result, the initial representations degrade very quickly. If any material needs to be retained, it must be moved from the sensory memory to the short-term memory (LaRue, 2013; Lawson et al., 2015). In the Pre-categorical Visual Storage, Sperling observed that the information was retained in a rather raw form prior to the linguistic material being categorised. This is unlike the Pre-categorical Acoustic Storage, which was assumed to survive longer than the visual one (see Roediger, Zaromb, & Lin, 2017, for a review). This may elucidate the reason for the superior immediate recall of auditorily presented items than those presented in the visual stream.

In summary, each modality has distinctive inherent idiosyncrasies (auditory and visual). As mentioned earlier, in visually presented language, for instance, recursive processing is available, and, as a result, the rate of input can be controlled by the reader. However, auditory presentation does not allow this, largely as a result of the nature of the rapid speech stream and its transience. In auditory presentation, prosodic and intonational cues are a prerequisite for learners to decipher the input stream, which, in turn, might impose more demands on processing resources, resulting in impaired auditory comprehension. Studies in L2 research (e.g., Murphy, 1997; Park, 2004; Wong, 2001) examining the effects of modality on L2 performance, including comprehension, have obtained a weight of evidence which supports the assumptions regarding the difference between modalities, namely, that L2 performance is more effective when such material is presented visually rather than auditorily. However, there is very limited research on the effects of modality on L2 learning itself (e.g., Sagarra & Abbuhl, 2013; Sydorenko, 2010) and the existing results were shown to be inconclusive. That is, some of the results contradicted the general claims about the differential effects of input modality in L2 performance. These results raise the question of whether the effects of input modality on L2 performance are different from those of L2 acquisition.

2.8   Working memory

As was indicated earlier, attentional capacity is constrained more in auditory than visual processing (Murphy, 1997; Wong, 2001). Learner's limited attentional resources imposes a WM capacity that provokes both the computation and storage of information. Accordingly, WM can be defined as a capacity system with a limited capacity of attentional resources responsible for the storage and simultaneous processing of information (including linguistic) for carrying out complex tasks such as reasoning, comprehension, and learning (Baddeley, 1999, 2007). Miyake & Shah (1999, p. 450) provided a particular definition:

"working memory is those mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition, including novel as well as familiar, skilled tasks".

This reveals that WM functions distinctly from short-term memory. Baddeley and Hitch (1974) argued that the latter is a unitary component or system and is merely concerned with the retention and repetition of information for short periods. Stated differently, WM is associated with the processing of information rather than just storing information, as is the case with short-term memory. An example of short-term memory is a Wi-Fi password that a receptionist in a hotel has given you orally for the first time. If somebody distracts you in the few seconds after you have heard the password, you will possibly forget the password and have to come back to the receptionist to ask for the password again. Here, it is important to maintain a distinction between short-term memory and long-term memory in that the latter holds information for a long time (e.g., months or years), which includes items such as a date of birth, the name of friends, and the fact that $5 + 3 = 8$ (Mazur, 2015).

In the early research on memory, memory was seen as either long-term memory or short-term memory, both of which stress the storage of information (Sakai, 2018). Later, 'short-term memory' was replaced with WM capacity (Baddeley & Hitch, 1974), where it is the mental aptitude that cognitive psychology and language research has focused on most. WM has been labelled "the hub of cognition", as it is the most influential cognitive mechanism of mental information processing (Haberlandt, 1997, p. 212). WM is also associated with intelligence, but the nature of this relationship is still a controversial issue (e.g., Ackerman, Beier, & Boyle, 2005; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008, see Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014, for a review).

According to Baddeley & Hitch (1974) and Gathercole & Baddeley (2014), WM capacity is a complex model comprising a central executive as a supervisory system that

controls and coordinates the operation of two underlying systems: the phonological loop and the visuospatial sketchpad. This multi-component system has received supportive evidence from both cognitive neuropsychology and brain-imaging studies that the phonological loop and the visuospatial sketchpad are separately processed by different memory systems and by distinct underlying mechanisms. The central executive is the most crucial part of the model and is responsible for control processes, such as monitoring responses, eliminating irrelevant information, switching retrieval plans, and retrieving information. It is also accountable for the allocation of attention or data to subsystems and relating them to the long-term memory.

The phonological loop has control of the process and the short-term retention of information encoded in a phonological pattern (auditory linguistic input). This loop is also in charge of the phonological representations of visually presented linguistic input. The phonological loop consists of two sub-units: the phonological store (inner ear) and a maintenance unit described as the 'articulatory rehearsal processes'. The former holds speech-based information of limited duration before it is lost to decay. The oral language automatically goes into the store as soon as it is encoded in a phonological format, while the visually presented language must be first encoded into a phonological form before it moves into the store. The articulatory rehearsal process hinders decay and displacement by reanimating the information of the phonological store through sub-vocal speech. It also has the function of encoding visually presented linguistic input into a phonological form (Baddeley, 1992; Goldstein, 2014).

The visual-spatial sketchpad (inner eye) as a second main component is concerned with the processing and storage of visual/imagery and spatial information. More specifically, it is "important for geographical orientation, and for planning spatial tasks" (Baddeley, 2002, p. 82). In other words, it keeps track of where people are with regard to other bodies when they move through the world (Baddeley, 2002). According to Allowy (2010), research in

brain imaging suggests that different parts of the human brain are able to recall verbal and visual-spatial information. It is further suggested that the lower area of the pre-frontal cortex manages verbal working memory while the higher area is associated with visual-spatial information.

The central executive (or, so called 'attentional controller') manages the action of both the phonological loop and the visuospatial sketch pad by allocating attention to particular parts of a task and switching focus from one part to another or dividing it between different tasks. Imagine you are a student in a biology class and your professor is giving an explanation of the digestive system, which you will be required to draw at the end, and meanwhile, a ceremony is being held beside the school. As your phonological loop is absorbing the professor's verbal explanation, your sketch pad is seeking to visualize an image for the digestive system. In this case, it is the executive centre's job to coordinate and merge the two types of information. The executive centre, further, is likely to turn your attention to the professor's speech and take you away from the speeches in the ceremony. Therefore, Unsworth, Heitz, Schrock, and Engle (2005) suggested that executive control is responsible for this relation where an individual retains attentional control in complicated tasks that need one to control interfering information.

2.8.1   The role of WM in L1 acquisition

There has been a burgeoning interest among cognitive psychologists to scrutinise the potential role of WM in first language processing and acquisition since the advent of the WM as a branch of knowledge in the early 1970s (Baddeley, 2003). However, researchers in Europe have distinct understandings of the WM–language relationship from those in the US (Andrade, 2002, the difference in the two concepts of WM discussed below is a summary of the discussion cited in Wen, 2015).

In most European countries, and in the UK specifically, researchers have mostly investigated the cognitive mechanisms pertaining to the phonological loop of WM (i.e., the unit in Baddeley's multicomponent model). More specifically, the WM–language connection has been investigated in terms of the role of the phonological loop in vocabulary and grammar learning among children (Baddeley, Gathercole, & Papagno, 1998). WM has been typically assessed by a storage-plus-rehearsal version of memory recall tests, such as non-word repetition span test (e.g., Gathercole, 2006; Gathercole, Willis, Baddeley, & Emslie, 1994). The results of these studies suggest that phonological WM (which implies a phonological short-term store and an articulatory rehearsal mechanism) produces an effect in acquiring novel phonological forms in both native and foreign language.

By contrast, researchers in the US (e.g., Patricia Carpenter, Nelson Cowan, Randall Engel, Andrew Conway, Michael Kane, Akira Miyake and Ellen Bialystok) have been more interested in the executive and control functions ascribed to the WM construct. Their research has focused on the potential role of the executive aspects of WM on language processing, particularly the sub-level processes that are involved in language comprehension. In assessing the WM, two more complex measures were used to assess the storage alongside processing, such as the reading span task constructed by Daneman and Carpenter (1980) and the domain-general operation span task advanced by Turner and Engle (1989), both of which are assumed to tax more of the executive functions of WM (e.g., information maintenance, updating, task switching, inhibitory control). A large number of studies (e.g., Acheson & MacDonald, 2009; Gibson, 1998; Hartsuiker & Barkhuysen, 2006; Miyake & Friedman, 1998) demonstrated a relationship between WM and language processing. In meta-analysis research carried out by Daneman and Merikle (1980), positive results were obtained to the point that WM has been correlated with reading comprehension processing. In a similar vein, studies (e.g., Acheson & MacDonald, 2009) have found a similar correlation between WM and language production.

Overall, according to Wen (2015), a close connection between WM and first language acquisition and processing can be established. In particular, based on the distinctive conceptions of WM, two aspects of WM have been found to be related to native language acquisition: the phonological loop and the central executive. The former, which is led by the European cognitive psychologists, has been demonstrated to be related to the first language acquisitional and developmental process in vocabulary and to be used in the long-term development of grammar, while the latter, which is guided by the US cognitive psychologists, has been found to be possibly involved in demanding and complex processes, specifically in language comprehension and production.

2.8.2 The role of WM in L2 language development

The critical data obtained from studies examining the relationship between WM and first language acquisition have led SLA researchers to the belief that the role WM plays in L2 learning is as crucial, if not even greater than in first language acquisition (Wen, 2012). This assumption is supported by salient central differences between L1 acquisition and L2 learning (Krashen, 1981; Meisel, 2011). L1 acquisition, for instance, mostly occurs implicitly and unconsciously. Its demands on WM are minimal, and more automatic processing takes place as a result of the immense exposure to language input 'repetition'. Such demands will only increase once such processing involves more demanding tasks, or the tasks to be processed are sufficiently inextricable (Wen, 2015). This is opposed to L2 learning, which mostly occurs explicitly and consciously, and thus more controlled processing is involved, presumably imposing greater demands on WM capacity (see Section 2.8.3 for more information on the difference between controlled and automatic processing).

In empirical studies investigating the relationship between the L2 and WM, the latter has been shown to play a role in a variety of L2 processes including syntactic processing (e.g., Dussias, 2003; C. N. Jackson & Dussias, 2009; Juffs, 2004) and comprehension (Chun

& Payne, 2004; Harrington & Sawyer, 1992; Osaka, Osaka, & Groner, 1993). Other studies, which are more closely related to the current study, have revealed that WM has a differential effect in L2 acquisition under different types of learning (e.g., Erlam, 2005; Martin & Ellis, 2012; Sanz, Lin, Lado, Stafford, & Bowden, 2014; Tagarelli, Mota, & Rebuschat, 2015). These studies reflect an expanding research trend interested in investigating both external (e.g., learning conditions) and internal cognition variables (i.e., interactionist perspective) for the purpose of elucidating the variation in the L2 acquisition development (Sanz, 2005, as cited in Sanz et al., 2014).

Erlam (2005), for instance, examined the relationship between aptitude including WM and the amount of learning resulting from three different instructional approaches: traditional deductive instruction (i.e., rule explanation, production practice, and explicit negative feedback), inductive instruction (i.e., production practice, input-based activities, and explicit negative feedback), and structured input instruction (i.e., rule explanation, input-based activities, and explicit negative feedback). Four tests, immediate and delayed, measured the instructional outcomes: listening and reading comprehension tests and written and oral production tests. The results showed the lack of any relationship between WM and the deductive group. Erlam (2005, p. 167) suggested that "deductive instruction that gives students opportunities to engage in language production minimizes any effect that individual differences in learner aptitude may have with respect to instructional outcomes". The WM measure, on the other hand, significantly correlated with the structured input instructions based on results from written production tests. The author concluded that learners with higher WM are better at processing explicit learning, which does not require participants to engage in production practice. However, it is worth noting that the WM capacity in this study was measured by a simple-span task, which lacked a processing component. This means it more

closely measures phonological short-term memory than WM capacity (Miyake & Friedman, 1998).

Some studies employed an artificial foreign language to control prior linguistic knowledge in order to examine how this L2 learning is affected by WM capacity. Martin and N. Ellis (2012), for instance, investigated how WM capacity and phonological short-term memory (PSTM) are correlated with implicit learning of grammar in an artificial foreign language. Singular forms of an artificial foreign language were learnt by 40 participants prior to exposure to plural forms in sentence contexts. The effects of implicit learning (i.e., use of plural markings and word order) were assessed using a generalisation test. The authors claimed that such a test measures not only participants' knowledge of structures but also participants' ability to generalize the forms to novel utterances. Participants also completed a non-word repetition task and a non-word recognition task as a measure of PSTM and a listening span task as a measure of WM. The results indicated correlations between PSTM and L2 grammar learning. Furthermore, WM was correlated with grammar learning, and the correlations between WM and grammar scores were consistently stronger than between PSTM and grammar. This difference in the strength of the correlation was expected because grammar learning is not restricted to memorizing items but also involves several mental processes related to abstract patterns in the sentences presented as input.

Tagarelli et al. (2015) examined whether WM capacity has a differential effect on L2 syntax acquisition under incidental and intentional conditions. A semi-artificial language comprising English lexicon and German syntax was used as stimulus material for the study. The language learning in both groups was assessed through an untimed GJT, including a confidence scale and the judgement base for each sentence. WM was measured by two complex span tasks: an operation–word span task and a letter–number ordering task. The results, overall, showed no significant correlation between performance on the untimed GJT

and both WM measures. Nevertheless, a statistical correlation was found on grammatical items for the explicit group. This led the authors to conclude that superior performance results from individuals with high WM when a learning condition is designed to boost the process of explicit learning. However, the authors acknowledged that the test used for language performance might have favoured explicit processes and suggested employing other measures that favour implicit unconscious processes (e.g., EIT or timed GJT).

2.8.3 WM effects on L2 acquisition and processing

In the acquisitional process of language, the significance of WM is evident in "storing the intermediate products of a reader's or listener's computations as he or she constructs and integrates ideas from the stream of successive words in a text or spoken discourse"(Just & Carpenter, 1992, p. 122). This implies that auditory/visual input processing, which involves various shared vital processes (for instance, word encoding, lexical access, syntactic and semantic analysis), is differently constrained by WM capacity, which provokes both computation and storage of information. The key constraint on WM capacity is imposed by limiting the amount of cognitive resources available for both maintaining and processing, which results in individual differences in WM capacity (Osaka & Osaka, 2007). That is, the more resources are available, the more WM operations are executed. Just and Carpenter (1992) and Y. Wu (1998) observed that when a task involves high cognitive demands, computation will slow down as a result of storage and processing necessities and, hence, the fidelity of recall of some partial results from WM processing may be negatively influenced. The principle here is assumed to be fairly similar to the processing that occurs in computers; when multiple processes are running simultaneously, task completion starts to be sluggish, and the machines can even crash.

Considering the distinctive idiosyncrasies of input between the modalities (in the case of auditory modality: variations in prosodic signals, rapid speed of input, and high degree of

transience; in visual modality: recursive processing, control over the input rate, stability in orthographic representation, see Section 2.7 for more details), the processing of visual input streams requires the processer to parse the input in the absence of elements related to prosody resulting in decreasing processing demands; consequently, WM capacity is subjected to fewer constraints. Similarly, the possibility of recursive processing and control on rate of input, as inherent idiosyncrasies of visual input, further reduces pressure on WM capacity, and as a result, large WM capacities may not significantly contribute to better performance in visual input tasks. In the case of auditory processing, a variety of cognitive processes must be at work in real-time to process the rapid input, and as a result, WM capacity is presumed to be a more critical variable in L2 auditory input than in visual input, making the former more cognitively demanding than the visual input (see Vandergrift & Goh, 2012, for a comprehensive review about auditory processing).

For successful auditory learning, greater automatization is required to allow more parallel cognitive processes to operate at a time, and more prompt availability for linguistic knowledge use is a prerequisite of this (see Buck, 1997). This is because distinct amounts of processing capacity are required to carry out various cognitive tasks (McLaughlin, Rossman, & McLeod, 1983). Before a task (e.g., access to L2 auditory or visual linguistic input) becomes automatic, it is learnt and practised through controlled processing (Schneider & Shiffrin, 2003). Controlled processing "occurs whenever you make a conscious decision to direct the processing activities of the brain" (Shiffrin & Schneider, 1977, cited in Williams, 2011, p.224). This means that this type of processing involves attentional control and therefore, exerts higher demands on WM due to a lack of parallel cognitive processing, which can, in turn, result in seriously impaired learning imposed by the constraints on the consumption of WM resources (Vandergrift & Goh, 2012). According to the L2 information processing theories developed by McLaughlin (1987, 1990) and Anderson (1983, 1985), as

cited in Bitchener & Ferris (2012), successful learning requires a shift from controlled processing towards automatic processing. Once L2 input is automatically accessed through practice or 'repeated activation', cognitive demands are reduced, and the connections between mental activities become more automatic. Automatized sequences are then stored in the long-term memory where they become available very rapidly whenever they are recalled. As a result, the demands exerted by input on WM capacity are minimal, as bundles of complex cognitive processes are activated concurrently. Put simply, WM resources can cope with the demands of information storage and processing, leading to less detrimental impacts on the amount of residual resources available for processing comprehension.

In the following figure, a model illustrates the underlying processes in WM when a new linguistic structure is learnt.
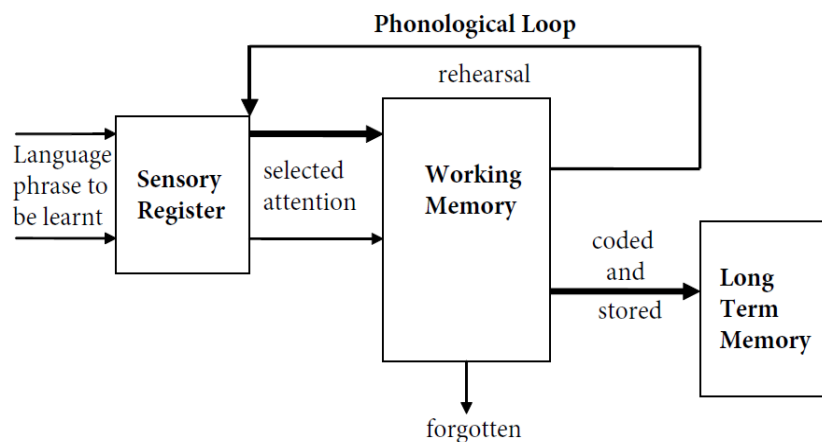
Figure 2. A model of how material can be learnt using the phonological loop in WM (taken from Randall (2007).

As indicated in the model, according to Randall (2007), the phonological loop provides the process of verbal (whether auditory or visual) repetition where the linguistic information remains automatized and stored away from decay and the possibility of loss. This will keep newly received information active while new information is interlaced with it to generate longer phrases or sentences. That is, if one maintains the connectionist perspective

that language is developed from the surface structure of the input, then repeated access to the stimuli is crucially important, as it strengthens the neural connections between nodes to particular memories.

To date, a large number of studies have investigated the relationship between WM capacity and L2 performance under a single pattern of input modality (auditory or visual), which has provided valuable results relating to how WM capacity interacts with the L2 performance depending on input modality. Research on the relationship between WM capacity and L2 reading has provided evidence that greater WM is favourable for L2 visual performance (e.g., Alptekin & Ercetin, 2010; Harrington & Sawyer, 1992; Leeser, 2007; Shin, Dronjic, & Park, 2019; Walter, 2004). Alptekin and Erçetin (2010), for example, examined the relationship of both L1 and L2 reading spans (as measures of L1 and L2 working memory) to L2 reading comprehension, where inferential and literal comprehensions were controlled. The former is deemed as a cognitively demanding task because it involves reading the whole text to elicit inferences while the latter is less demanding because it relies on the text-based level. The authors claimed that WM capacity might have a distinct effect according to the type of reading comprehension involved, that is, either literal or inferential. The former involves deeply analytic reading, where novel knowledge based on reasoning and evidence is produced from textual content and where comprehension is presumed to induce more resource-demanding control processes. The latter requires text-based information processing where conceptual and linguistic operations run at a lower level than inferential comprehension. Literal comprehension is presumed to involve fewer resource-demanding control processes than inferential comprehension. In this study, 43 Turkish university students completed three tasks: L1 reading span task (in Turkish), an L2 reading span task (in English), and an L2 reading comprehension test using multiple-choice items based on an

English short story. The results showed no significant difference between L1 and L2 WM, and the WM capacity was found to be correlated with L2 inferential comprehension.

In the same vein, Shin et al. (2019) investigated whether there is any interaction between WM and background knowledge in L2 reading comprehension. A total of 74 Korean L2 learners performed three tasks: a reading span task, a vocabulary size test, and a C-test (a measure of general language proficiency). On a different day, the participants were exposed to two background texts on their own, and they then completed a TOEFL reading test the following day, where the test included texts with and without any relevant background knowledge. The results indicated that the participants with higher WM gained more advantage from the provision of background knowledge, which resulted in superior reading comprehension, than those with low WM. The authors concluded that the relationship between WM capacity and L2 reading comprehension can be affected by learners' prior knowledge and that the level of linguistic knowledge and knowledge on the topic might have played a crucial role in L2 readers' efficient use of their WM capacity.

Jeon and Yamashita (2014) conducted a meta-analysis of 58 previous studies examining the relationships between L2 reading comprehension and 10 linguistic and cognitive components, including L2 decoding, L2 grammar knowledge, L2 vocabulary knowledge, L1 reading comprehension, L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, L2 listening comprehension, WM, and metacognition. The results revealed that L2 grammar knowledge and L2 vocabulary knowledge related strongly to L2 reading comprehension, whereas WM reflected a weaker predictor ($r = 0.42$).

However, research on the relationship between WM and L2 auditory performance is sparse compared to that on L2 visual performance, where the findings of existing studies (e.g., Andringa et al., 2012; Kormos & Safar, 2008; Miyake & Friedman, 1998; Vandergrift

& Baker, 2015) have been inconclusive. Vandergrift and Baker (2015), for instance, carried out an exploratory study investigating the degree to which cognitive L2 learner variables, including WM, might predict success in L2 listening. A total of 157 learners of French, with a mean age of 13, were recruited for the study. The participants' first language was considered English based on if they were either born in Canada or if they had completed all their schooling in English. They were sampled from three different groups based on the calendar year the students joined the program: 2008 ($n = 47$), 2009 ($n = 49$), and 2010 ($n = 61$). The L2 listening was measured via the French listening comprehension test, which consisted of 28 multiple-choice items. To test WM capacity, the backward digit recall and the non-word list recall tests were run. The scores of the two tests were merged to produce one score to measure WM capacity. The results indicated that WM may be tied to L2 listening comprehension, but the relationship was found to be significant only for the first cohort. The non-correlation result might be due to the use of the non-word list recall task, which is more useful in measuring the phonological short-term memory than WM capacity.

Andringa et al. (2012) addressed the same question but with native and L2 listening ability. In total, 121 native Dutch speakers and 113 non-native speakers of Dutch (they were from 35 different first languages) were tested on the various linguistic skills (e.g., discourse comprehension, vocabulary, semantic processing, grammatical processing, and self-paced listening) that are presumed to underlie listening comprehension. WM capacity was measured using four-digit span tasks (forward and backward visual and forward and backward auditory) and one non-word recognition task. When WM was expressed as a single factor including all five tasks, the results revealed that the relationship was not strong enough to explain the unique variance in L1 and L2 listening comprehension. However, a crucial limitation of the study concerns the digit forward span tasks used as additional measures of

WM capacity. Such tasks lack a processing component; that is, they more closely measure phonological short-term memory than WM capacity.

In a qualitative study, Goh (2000) analysed learners' self-reports using diaries, interviews, and immediate retrospective verbalisations with 40 Chinese learners of English to explore a cognitive perspective on the comprehension issues of L2 listeners. Ten issues were identified that the learners encountered during the L2 listening. Five challenges were related to word recognition and attention failure during perceptual processing, while others were linked to ineffective parsing and failure to use the mental representations of incoming language. The most frequently mentioned issue was "quickly forget what is heard" (p. 60). This issue was ascribed by the author to the learners' limited WM capacity, where it is demanding for L2 learners to process and store the just-heard material and to attend to the new material simultaneously.

Correlation between WM capacity and L2 auditory performance may depend on other factors which are taken into account. For instance, Brunfaut and Révész (2015) explored the relationship between L2 listening and listeners' WM. A total of 93 L2 learners of English performed 30 multiple-choice listening items and an English proficiency test (Pearson Test of English Academic). Two WM measures were administered: auditory forward and backward digit span tasks. The analysis of data revealed moderate correlations between WM measures and the overall listening scores of the English proficiency test, but no significant correlation was obtained between WM and the listening scores of the 30-item passage completion multiple-choice test. The authors ascribed this result to the differences in the two test types: the English proficiency test required L2 listeners' global comprehension (i.e., listening for main ideas) while the multiple-choice test was designed to measure L2 listeners' local comprehension (i.e., listening for specific details). They suggested that listening tests that involve local comprehension may relate to WM capacity.

Although previous studies have sought to examine how WM capacity is correlated with learners' performance on a single modality, as shown in the studies above, research that applies two receptive modalities (i.e., auditory vs. visual input) is very limited. Kormos and Safar (2008), for example, investigated the relationship between WM capacity and performance in an end-of-year reading, writing, listening, speaking, and use of English test (vocabulary and grammar). A total of 121 secondary school students aged 15–16 participated in an intensive English language training in Hungary. Of the 121 participants, 50 completed a backward digit span test to measure their WM capacity. At the end of the program, a Cambridge First Certificate Exam was administered to all participants. The findings revealed that WM correlated very highly with the overall scores on the language exam, including all separate skills, with the exception of writing. More specifically, the strength of correlation between WM capacity and listening was similar to the strength of correlation between WM capacity and reading.

This literature review on WM and L2 input modality has, so far, primarily concentrated on the L2 performance achieved by the learner during language tests. Another crucial issue should be highlighted, that is, how WM capacity affects the process of L2 learning itself. This specifically refers to when such learning is delivered under two distinct input modalities, namely, auditory versus visual. Studies examining this relationship (i.e., the relationship between WM and the amount of learning accrued from two different modalities 'auditory vs visual') are very limited; indeed, to the best of the author's knowledge, there is only one study (e.g., Sagarra & Abbuhl, 2013) that has addressed this relationship directly. Sagarra and Abbuhl (2013) investigated the effect of WM in the performance of gender and number agreement in L2 Spanish through feedback modality (visual vs. auditory) where participants were required to produce adjectives correctly to complete sentences. Participants were assigned to one of six experimental groups and exposed to computer-delivered feedback

(see Section 2.7 for more details about the six types of experiments). The control group did not receive any feedback. The effects of learning were assessed by three written posttests as well as two oral interactional tasks. The results indicated that participants' performance on immediate and delayed tests (written and oral) were found to correlate with WM when participants received an auditory recast (both enhanced and unenhanced). The authors concluded that the lack of effect of WM on the visual recast groups may be relevant to processing demands. That is, the processing demands placed on WM are fairly minimal when visual input is processed.

In summary, research on the relationship between WM capacity and L2 input modality provided evidence that WM capacity plays a factor in L2 visual performance. However, research on the relationship between WM and L2 auditory performance has not received significant attention, and the existing results have been inconclusive. More importantly, only very limited research has addressed the relationship between WM capacity and the amount of learning accrued from two different input modalities (auditory and visual). Only one study (e.g., Sagarra & Abbuhl, 2013), to the best of the author's knowledge, has addressed this relationship and found a correlation between WM and auditory learning but not with visual learning. Compared to previous studies, which indicated that WM mediated the effects of visual modality on L2 performance while the results for the effects of auditory input are inconclusive, Sagarra and Abbuhl's (2013) study appears to be the first study where the results are in line with the general assumption that WM capacity is a more critical variable in L2 auditory input than in visual input. In light of the conflicting findings between Sagarra and Abbuhl's (2013) study (which focused on L2 learning) and the other studies (which focused on L2 performance), further research exploring whether the effects on language development of input modality during learning are mediated by WM, as in the

present investigation, can enhance our understanding of the relationships between WM and L2 learning accrued from auditory versus visual modality.

2.9 Summary of literature review

Overall, five salient points can be concisely drawn from the literature reviewed above as a navigation base for the research design in the current investigation.

First, automaticity as fluent language processing is indicative of implicit knowledge acquisition. In L2 research, the principle is how grammatical knowledge can become available for use in automatic processing reflected in fast processing, with limited or no explicit monitoring. This is irrespective of the distinct theoretical views regarding whether explicit knowledge is proceduralised to become automatized knowledge, is 'functionally equivalent' to implicit knowledge, or is simply implicit knowledge acquired implicitly. Similarly, the understanding of implicit knowledge in the present study did not take the notion of awareness into account. This is because the stimuli material was generated using natural language (i.e., English) with which participants were already familiar. This concept is followed in several previous research studies examining the effect of specific learning mechanisms on the development of L2 implicit knowledge acquisition (e.g., Akakura, 2012; R. Ellis et al., 2009; Erlam et al., 2009; Reinders & Ellis, 2009).

Second, the elicited imitation task (EIT) and timed grammaticality judgement task (timed GJT) reflect underlying automatic processing in a similar way, both of which presumably are possible measures of L2 implicit grammatical knowledge. For the EIT, to ensure that such a measure taps into stronger implicit knowledge, several criteria should be taken into consideration when designing a version of EIT: i) attending to the semantic meaning of an utterance rather than its syntactic form so that the likelihood of rote repetition is reduced, ii) including ungrammatical items to inspect if the syntactic anomalies are repeated verbatim or spontaneously rectified, iii) controlling the time lag between the

auditory stimulus and the oral imitation so that the possibility of greater access to explicit grammatical knowledge is minimised, and iv) due to the requirement of 'oral repetition' in the EIT, the capacity for storing information in short-term memory should not greatly affect EIT performance. Therefore, a digit span task, as a measure of phonological short-term memory, was deployed to control the effects of short-term memory capacity.

In the case of a timed GJT, on the other hand, the most influential factor in terms of inducing learners to access implicit knowledge is the amount of time participants are given to make judgements. Such a task requires a participant to judge under time constraints, from the first impression, whether a visually presented sentence is grammatically acceptable or not. Rapid linguistic response to the GJT sentences reflects a high degree of automaticity of the acquired knowledge.

In light of the above, both an EIT and a timed GJT were used as tests to observe participants' development of L2 implicit grammatical knowledge. The rational for deploying both was, first, to find out to what extent the results obtained from the EIT were consistent (i.e., degree of reliability) with another test (i.e., timed GJT) tapping into the same type of knowledge (i.e., implicit knowledge), and second, and more importantly, using a test presented auditorily (i.e., EIT) and a test presented visually (i.e., GJT) was necessary in the present investigation to prevent any prejudice in favour of one input modality over another during learning.

Third, two types of learning conditions are presumed to promote the acquisition of L2 implicit knowledge: focus on form and oral output. The former allows for form-meaning mapping where linking grammatical forms to the meanings the learners comprehend from the input is most likely to engender the acquisition of implicit knowledge (Doughty, 2001; R. Ellis, 2006; Long, 1991). Although a limited number of empirical studies have addressed the effects of focus on form on implicit knowledge acquisition (e.g., Pietrzykowska, 2011;

Reinders & Ellis, 2009), their results can be interpreted as indicating that focus on form is able to promote learners' linguistic competence (i.e., implicit knowledge).

The latter (i.e., oral output) is premised on cognitive interactionist theories, which suggest that input in conjunction with language output triggers the internal mental processing required in acquisition, resulting in speeding up the L2 acquisition. Rebuschat and Williams (2012) and Erlam et al. (2009) offered evidence to support the hypothesis that oral output can have a facilitative effect on implicit knowledge acquisition. The former found the oral output practice fosters automaticity based on the use of EIT, while the latter found that oral output practice produced a greater learning effect in terms of unconscious implicit knowledge (i.e., at least some of the learning was unaware learning).

In light of the three-point discussion above, the present study employed focus on form as training in both modality groups (auditory vs. visual) where participants in each group were also engaged in oral output practice. To ensure that participants had comprehensive, meaningful training on the target constructions, they were also be exposed to a deductive learning task (i.e., an explanation of a grammatical rule is provided first, followed by concrete examples of language use). Putting all training tasks together, participants in each modality group were first exposed to the focus on form task and then subjected to a deductive learning task where they were engaged in oral output practice in each task (see Section 3.5.1 for more information about the training stage). This task order allowed a participant to adapt to the use of the target structures in a context in which the primary focus was on the meaning of the message (i.e., connecting the grammatical forms with their meanings). It also gave the participant at least a sporting chance of acquiring the target structures incidentally before their exposure to explicit learning. It must be noted that the main goal of the present project lies in identifying differences in the acquisition of implicit knowledge between the visual input and auditory input modality groups that took part in the same training. Therefore,

identifying the differences in performance between focus on form and deductive learning tasks is unfeasible.

Fourth, although comprehension as a general construct has some basic principles shared by auditory and visual processing input, the former has distinctive inherent idiosyncrasies which impose heavy demands on processing resources, leading to detrimental effects on L2 performance. Despite sufficient empirical evidence to support this hypothesis, research investigating the effects of input modality on L2 learning itself is limited (e.g., Sagarra and Abbuhl, 2013; Sydorenko, 2010); in addition, the existing results have been inconclusive and some have revealed counter evidence (i.e., an advantageous bias towards auditory input). Thus, in light of these conflicting results in the literature, further research investigating the relative effects of input modality (auditory vs. visual) on the L2 acquisition itself through the use of implicit knowledge measures (i.e., an EIT and a timed GJT), as in the present study, can provide more critical data regarding the effects of input modality on L2 acquisition.

Fifth, previous research has provided evidence that WM capacity is a critical variable in L2 visual performance. However, research on the relationship between WM and L2 auditory performance has not received significant attention, and the existing results have proved to be inconclusive. More importantly, very limited research has addressed the relationship between WM capacity and the amount of learning accrued from two different input modalities (auditory vs. visual). Only one study (e.g., Sagarra & Abbuhl, 2013) has addressed this relationship and provided results supporting the general claims (i.e., WM capacity is a more critical variable in L2 auditory input than in visual input). Despite the successful work in Sagarra and Abbuhl's (2013) study in terms of WM capacity, the substantial challenge remains regarding whether WM capacity still interacts with the amount

of learning accrued from each input modality when such this amount of learning is gauged through implicit knowledge measures.

Chapter 3: The Present study

The current study builds on previous research literature and aims to scrutinise two factors which might affect the development of L2 implicit knowledge: modality of language learning and WM capacity. The studies described above revealed evidence to indicate that the modality of input presentation influences L2 performance, resulting in superior performance with visually presented language. This superior performance is assumed to result from distinctive idiosyncrasies inherent in the nature of visual processing (e.g., immutability in orthographic representation, recursive processing and control over the input rate) generating the deeper processing of linguistic information. However, there is a relative lack of SLA studies comparing the effects of input modality on L2 learning itself (e.g., Sagarra & Abbuhl, 2013; Sydorenko, 2010); the existing results were inconclusive, and some revealed counter-evidence (i.e., an advantageous bias towards auditory input). More specifically, it is not yet clear which input modality results in a more successful acquisition of implicit knowledge.

Therefore, the present study furthers research on this topic by investigating whether there is a difference in the development of L2 acquisition depending on the input modality used during the learning process. However, the present investigation differs from the previous studies, as the language development (accrued from auditory vs. visual input modalities) in the current investigation is gauged by the measures of implicit knowledge (i.e., an EIT and a timed GJT) following the assumption that such measures would provide more critical data regarding L2 acquisition. That is, they are more likely to indicate whether the L2 developing system is predisposed to automating (i.e., successful emergence of implicit knowledge) via auditory or visual input based on language learning.

The second factor which might affect the process of acquiring implicit knowledge is how the effects of WM capacity are related to the amount of learning accrued from two distinct modalities (auditory vs. visual). Research examining the relationship between WM

capacity and the amount of learning accrued from two different input modalities (auditory vs. visual) is very limited. Only one study (Sagarra & Abbuhl, 2013) has addressed this relationship and elicited results that are in line with the general claims (i.e., WM capacity is a more critical variable in L2 auditory input than in visual input). This is unlike research on the relationship between WM capacity and the input modality of a language task where results, generally, have provided fairly contradictory data; visual input modality is related to WM capacity more than it is to auditory input modality. Despite the interesting work in Sagarra and Abbuhl's (2013) study in terms of WM capacity, the question of whether WM capacity predicts how implicit grammatical knowledge is acquired differently with exposure to two different input modalities has not yet been investigated.

Despite the proven success of previous experimental work, the absence of studies investigating the role of the modality of stimulus presentation in the success of implicit knowledge acquisition and whether WM capacity interacts with this modality is apparent. Therefore, the current investigation aims to fill two gaps in the knowledge in this field: first, by attempting to distinguish the effects of auditory/visual input on the acquisition of L2 implicit knowledge, and second, by exploring whether WM capacity mediates the acquisition of implicit grammatical knowledge in the case of auditory vs. visual input modality.

To fulfil the aim of the present investigation, two experiments were performed. Experiment 1 involved Chinese L1 speakers ($n = 77$), while Experiment 2 was identical to Experiment 1 but involved Arabic L1 speakers ($n = 37$). The purpose of carrying out Experiment 2 was to verify the findings of Experiment 1 in a different language group.

3.1 Hypotheses and research questions

The experiments were conducted to investigate the following hypotheses:

H_0: There is no significant difference in the effects of auditory and visual input modalities on the acquisition of implicit grammatical knowledge.

H₁: There is a significant difference in the effects of auditory and visual input modalities on the acquisition of implicit grammatical knowledge.

H₀: WM capacity does not mediate the acquisition of implicit grammatical knowledge in the case of auditory vs. visual input modality.

H₁: WM capacity mediates the acquisition of implicit grammatical knowledge in the case of auditory vs. visual input modality.

The hypotheses above can be articulated in the following research questions:

1. Which of the two input modalities (auditory vs. visual) can result in more successful acquisition of implicit grammatical knowledge?

2. To what extent does WM capacity mediate the acquisition of implicit grammatical knowledge in the case of auditory vs. visual input modality?

3.2   Research approach

A quantitative approach using a pretest/posttest design that draws on numerical data, followed by statistical analysis, is the most appropriate to employ in the current study. Such approach allows for establishing cause–effect relationships between the variables (Muijs, 2010, i.e., the effects of input modality on the development of implicit grammatical knowledge). Thus, confounding variables were reduced, and more sensible interpretations of the findings can be obtained. Quantitative research requires instruments and data collection with as high a level of objectivity as possible and with a decrease in subjective interpretations (Mertler, 2019). Therefore, the researcher in the current study attempted to adopt a neutral position to eliminate any bias which might affect the results of the current investigation.

A pre-/posttest design allows any changes that may have occurred in either group to be observed by comparing the pre- and posttest performance for each group (auditory or visual). Not only can the pre- and posttest performances be compared within the two groups, but the groups can also be compared from the pre- to the posttest (S. L. Jackson, 2014). That

is, if the visual modality has some effect, then the visual group should have a greater improvement from pre- to posttest than that of the auditory group, and vice versa. However, there is a threat to this design's internal validity, namely, pretests could serve as a learning experience for participants. Consequently, the effects of a treatment might be biased, though this threat can be controlled by using another version of measures in the posttest, as is the case in the current study (see Section 3.5.2 for more information about the versions of measures). In essence, a posttest-only design is likely to provide more accurate data but is unfeasible in the current study without completely controlling for the participants' prior knowledge of the stimulus material by using an artificial or semi-artificial language that is novel to the participants. Instead, the current thesis employed a natural language to generate the stimulus material so that any implications resulting from the use of a natural language can be more practical for teaching and learning than if using an artificial language.

3.3   L2 Target syntax

Control over prior knowledge of stimulus material poses a challenge in SLA-based experimental research studies. Higher internal validity and accuracy of data can be elicited from an experiment when the stimuli are novel (e.g., artificial language) or at least infrequent and uncommon to participants. Similarly, the fewer target stimulus structures there are, the greater the possibility of gaining many data points per structure if the total number of stimuli in the experiments is fixed. This opportunity can maximise the validity of certainty whether the independent variable (visual or auditory) is the cause of the change observed in the dependent variable (acquired automatic knowledge). It will further reduce the potential threats to internal validity, as extraneous variables (e.g., factors related to linguistic structure) might facilitate or inhibit the task (Bley-Vroman & Chaudron, 1994). Hence, the stimulus material in the present study was generated by three complex structures which are known to be problematic for L2 learners: i) negative adverbs with an inverted subject and auxiliary

verb (e.g., *Rarely have I seen such a huge building*), ii) the past counterfactual conditional with subject–auxiliary inversion (e.g., *Had he known the answer, he would have told her)*, and iii) canonical tag questions (e.g., *David was responsible for the burglary, wasn't he?*).

The properties of the three structures will be discussed separately below in terms of five dimensions for the purpose of determining the level of difficulty for each structure. These dimensions include processability, regularity, input frequency, perceptual saliency, and functional value, where they are explained in more detail in the section dealing with the first structure. The dimensions were derived from N. Ellis's difficulty criteria (2006) and Pienemann's processing theory (1998, as cited in R. Ellis, 2009b).

3.3.1   Negative adverbs

Negative adverbs (NA structure) are adverbial structures that result in the inversion of subjects and an auxiliary verb: The negative adverbial is positioned first, an auxiliary verb follows, and the subject of the sentence comes next. This occurs in a sentence such as '*Seldom had James done his job in time'*. Other negative adverbs include 'never', 'rarely', 'seldom', and 'hardly'. Swan (1988) stated that negative adverbs with an inverted subject and verb are employed to make the sentence more emphatic and formal, and very often appear in literary works. Therefore, this grammatical feature is acquired later in L2 acquisition (i.e., better understood by advanced learners (Reinders & Ellis, 2009).

3.3.1.1 Processability

Processability refers to the extent to which an L2 learner is able to process a structure (via ordering) to construct a grammatically correct order of form. Pienemann (1998, p. 46) stated that "the psychological complexity of a structure is dependent on the degree of reordering and rearrangement of linguistic material involved in the process of mapping underlying semantics onto surface forms". In the case of the NA structure, it is considered

difficult to process because the word order governing the sentence must be rearranged to accommodate a syntactically correct form.

3.3.1.2 Regularity

Regularity refers to the extent to which the grammatical feature conforms to other grammatical patterns (i.e., the number of grammatical cases applying to the underlying syntax governing the target grammatical feature). Regular features are hypothesised to be easier to learn than irregular features. In the case of the NA structure, only NAs, specifically when they occur at the beginning of the sentence, generate subject-auxiliary word order inversion (e.g., *Rarely have I seen huge buildings in this city*), while the other adverbs, such as adverbs of time, place, and manner, generally assume a normal subject–verb word order (i.e., subject/verb/object, e.g., *Occasionally I have seen huge buildings in this city*). This suggests that the NA structure is an irregular feature, and thus it is difficult to acquire.

3.3.1.3 Input frequency

Frequency refers to the extent to which the grammatical feature occurs in the input. The underlying assumption emanates from connectionist theories of language learning where the linguistic structures that occur frequently in the input can be easily acquired as opposed to structures that occur rarely. In the case of the NA structure, it is comparatively infrequent and is restricted to formal situations, such as literary writing, as affirmed by an analysis of the British National Corpus (which has millions of entries) in which, "frequency ranged from 276 occurrences for 'not only was' to 3 for 'seldom do' and fewer for a range of other adverb/auxiliary combinations" (Reinders & Ellis, 2009, p. 290).

3.3.1.4 Perceptual saliency

Saliency refers to the extent to which it is straightforward to identify the grammatical feature in the input. Syntactic items are more salient when they occur in anchor positions

(i.e., at the beginning or end of a sentence). Via artificial grammar learning experiments, Chan (1992, cited in Allwood et al., 2000) observed that successful language acquisition is more likely to take place when the target syntactic items occur in sentence-initial or sentence-final position. In the case of NA structure, the element (the adverbial) that provokes the inversion is in the sentence-initial position, but the resulting inversion is not (Reinders, 2010) or is less salient because the auxiliary is usually unstressed (Reinders & Ellis, 2009). Long (1996, as cited in Reinders, 2010, p. 99) stated that a "subject-auxiliary inversion after proposed negative adverbials (*seldom have I seen…*) and uses of whom, are too rare or perceptually nonsalient".

3.3.1.5 Functional value

Functional value refers to the extent to which the grammatical feature maps onto a clear, distinct function. Grammatical structures typically possess discoursal, semantic, or pragmatic functions. However, some structures (e.g., 3rd person -s) are entirely redundant. It is hypothesised that a structure which has a single function and is typically non-redundant will be easier to acquire than structures that possess multiple functions or are redundant. In the case of the NA structure, although negative adverbs (e.g., rarely, hardly, and seldom) might be straightforwardly identifiable, they do not indicate complete opposition as in the case of 'not' or 'never'. For example, the adverb 'rarely' means 'not often'; although its meaning and function reflect negativity, it simultaneously bears some indication that something is present or takes place despite its infrequency (see Reinders, 2010, for a review). This semantic complexity, as Reinders (2010) suggested, makes it harder for an L2 learner to extract the syntactic rule governing this structure: "In many cases, perhaps, no rule is formed and only individual items are remembered as causing inversion" (p. 100). Altogether, the negative denotation inherent in the adverb is lexically represented, and the subject-auxiliary inversion is redundant, and thus the NA structure is considered difficult to acquire.

3.3.2   Counterfactual conditional

Conditionals (CON structure) entail two clauses where the dependent clause expresses a condition required for fulfilling the independent clause. In formal English, the adverbial subordinator *if* can be omitted, which triggers a subject-auxiliary inversion, as in the sentence, '*Had you worked hard in your bachelor's studies, you would have gained higher grades*'. Generally, counterfactual conditionals remain the most complicated type of all conditional clauses (i.e., real conditional, hypothetical conditional, or counterfactual conditional), both syntactically and semantically (Berent, 1985). Berent (1985) performed an experimental study on 55 adult learners of English to investigate which of the conditionals is more difficult for production and comprehension. The results of the study showed that although the counterfactual conditionals were easy to understand, they were the most complex type to produce. This finding presumably explains the complicated syntax governing the combination of the two clauses in the past counterfactual conditionals.

3.3.2.1 Processability

The syntactic structure of counterfactual conditionals with subject-operator inversion involves a subordinate clause composed using the combination of the operator 'had' + subject + a past perfect in the conditional clause (i.e., the dependent clause) and a model perfect (subject + would + have + past participle of the verb) in the main clause (i.e., the independent clause). Such a structure is considered difficult to process because, first, it mainly requires two clauses, and second, the adverbial subordinator *if* needs to be omitted, which triggers inverting the operator *had* with the subject in order to create a syntactically correct form.

3.3.2.2 Regularity

The counterfactual conditionals with the initial subordinator *had* can be considered to fail to meet the regularity criterion because such a case involves subject-auxiliary inversion,

unlike the counterfactual conditional, in which the adverbial subordinator *if* generally adheres to normal subject-auxiliary word order.

3.3.2.3 Input frequency

Counterfactual conditionals are relatively infrequent, as affirmed by analysis of the conditionals in the corpora. Hwang (1979, as cited in R. W. Norris, 2003), for instance, found that counterfactual conditionals appeared in 10 (3.8%) out of 266 conditionals in the spoken corpora, while they appeared in 31 (3.3%) out of 948 conditionals in the written corpora. Flucher (1991) conducted a corpus study of clauses using 'if'. The study indicated that the three conditional types accounted for 61 (20.4%) out of 299 occurrences of 'if' forms. The occurrence of a combination of 'if' + a real conditional or 'if' + a hypothetical conditional appeared 135 (45.15%) times in total. However, the counterfactual conditionals showed only 9 (3%) of the 299 conditionals. Thus, if the counterfactual conditional (if+ past perfect …) occurs infrequently based on the analysis of the two studies' results, it is entirely possible — if not probable — that the occurrence of counterfactual conditionals with subject-operator inversion (e.g., *Had they played football, they would have won*) is very rare, specifically amongst the target sample in the present study.

3.3.2.4 Perceptual saliency

Although the subordinator *had* that provoked the inversion is in the sentence-initial position, the resulting inversion is less salient. This finding can be supported by Ellis's classification (2009) for hypothetical conditionals (e.g., *If David fell, he would hurt himself*) where he classified it as having low saliency.

3.3.2.5 Function value

The dependent clause *'If you had worked hard in your bachelor's studies'* refers to a counterfactual past event that contradicts the state of events in the real world. The condition

is presented to express extreme remoteness in both time and possibility (i.e., counter to fact, it did not happen, Yule, 1998). Furthermore, the CON structure bears the underlying suggestion that the speaker expresses an excuse, regret, or relief. It is crucial to note that the precise meaning of conditional clauses sometimes depends on the types of modal verbs included in them. In hypothetical conditionals, for instance, the condition 'if' can be weakened by using the modal *should*, the verb *happen to*, or both together — *should happen to* — rather than the past tense in the dependent clause, as in the sentence '*If he happened to have time, he would submit his paper early*' This change weakens the negative sense of the '*if*' clause and strengthens the possibility of the result. More importantly, however, this strengthening of the possibility does not arise in counterfactual conditionals because the 'if' clause is strongly negated, and the condition remains impossible (see Norris, 2003, for a review).

3.3.3 Canonical tag questions

A tag question (TQ structure) is a short question which occurs immediately following a declarative sentence or an imperative statement in a conversation, or in written representations of speech (e.g., *David was responsible for the burglary, wasn't he?*). It is used when a speaker formulates a hypothesis about a specific attitude or event and seeks confirmation. The rising and falling tone plays a major role in its function (Downing & Locke, 2006). Canonical tag questions in the English language are complex and have a form that is unique to the English language.

3.3.3.1 Processability

The structure as a whole requires the processing of two different clauses: The first is a declarative statement representing a stem, and bears an assertive property, while the second, which is attached to the stem, is interrogative (a tag) represented as a short question form. More importantly, the appropriate TQ form is entirely governed by the independent clause.

That is, a subject pronoun and an auxiliary verb must stem from the declarative sentence (the first clause), and then a subject-auxiliary must be reverted to form a tag question that has a tendency towards affirmation or negation. This means gender, number, and tense must also be taken into account in order to construct a syntactically correct tag question. This syntactic form of the value (affirmative or negative) varies based on the response a questioner expects from the interlocutor. If the expected response is 'yes', then an affirmative statement and a negative tag must be used. However, the two values can be reversed when a negative response is expected (i.e., a negative statement plus an affirmative tag). In spoken language, native speakers of English implement these rules automatically and unconsciously. However, L2 learners may find it difficult to take gender, number, and tense into account while speaking in order to meet grammatical requirements.

3.3.3.2 Regularity

Tag questions are considered to meet the regularity criterion. The subject-auxiliary inversion is large in scope because the inversion is a prerequisite for composing a tag question in English. Such an inversion is also high in reliability as a result of a large application of the normal inversion, which must agree with the tense, aspect, and modality of the verb in the declarative sentence (e.g., 'I am right, aren't I?', not 'amn't I').

3.3.3.3 Input frequency

Tag questions are frequently used in everyday communication. Based on the British component of the International Corpus of English, of 754 instances, only 15 tag questions in the written texts were observed, and the remaining 739 instances were included amongst the spoken texts. This analysis supports the general assumption that tag questions are predominant in spoken language (J.-B. Kim & Ann, 2011).

3.3.3.4 Perceptual saliency

Tag questions are perceptually salient, as they occur at the end of a sentence.

3.3.3.5 Functional value

The rising and the falling tone within the tag plays a major role in its semantic function. A rising tone in the tag indicates doubt; the questioner is requesting more information. However, if the intonation is falling, it expresses greater certainty, so that the meaning of the tag is requesting confirmation (for example, 'I'm asking you to confirm this') and simply seeks agreement (Downing & Locke, 2006).

Table 1

Summarising the Determinants of the Relative Grammatical Complexity of the Three Structures

	Criteria of difficulty				
Rule	Frequency	Saliency	Regularity	Processing	Function
NA	Low	Low	Irregular	Difficult	Difficult
TQ	High	High	Regular	Difficult	Difficult
CON	Low	Low	Irregular	Difficult	Difficult

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

To summarise, the present study employed experimental material that is complex and known to be problematic for L2 learners. Three syntactic structures in English were employed as stimulus materials: i) negative adverbs with an inverted subject and auxiliary verb (e.g., *Rarely have I seen such a huge building*), where the linguistic focus is on three adverbial structures that trigger the inversion of subject and auxiliary verb: *seldom, hardly, and rarely;* ii) the past counterfactual conditional with subject–operator inversion (e.g., *Had he known the answer, he would have told her)*, and iii) canonical tag questions (e.g., *David*

was responsible for the burglary, wasn't he?). According to the aforementioned dimensions (frequency, saliency, regularity, processing, and Functional value), TQ appears to be less difficult to automatize than NA and CON due to its high frequency and its regularity. Thus, L2 learners with IELTS scores of 5.5, 6.0, and 6.5, as was the case in the present study, will have some explicit knowledge of this structure. However, the aim of the current investigation was not to examine which input modality promotes the learning of a completely new structure but, instead, to determine which modality fosters learners' ability to attain greater control over a structure that can become available for use in automatic processing (i.e., can acquire the implicit knowledge of the structure).

3.4   Research design

This project adhered to a pretest/posttest design to examine the differences in the development of automatized linguistic knowledge in two experimental groups who had been subjected to training in the auditory versus visual modality. At first, all qualified participants (whose IELTS scores fell in the targeted range) were randomly assigned to one of the two modality groups. The data collection involved two main experimental sessions: the first session included i) an administration of two language tests, an EIT and a timed GJT, both of which measure implicit grammatical knowledge, and ii) an administration of two memory measures, an OSPAN task as a measure of WM capacity, and a digit span (DS) task as a measure of phonological short-term memory. In the second session, the intervention (consisting of incidental and explicit learning tasks) was introduced to the groups. The participants in each group received an identical amount of exposure. Finally, a posttest similar to the pre-test was immediately administered. The second session was accomplished no less than 2 days and no more than 7 days following the first session. Each session took approximately 1 hour. The experiment was conducted in two locations: the School of

Experimental Psychology and the School of Education at the University of Bristol (in some cases in Experiment 2, the experiment was conducted in participants' preferred places).
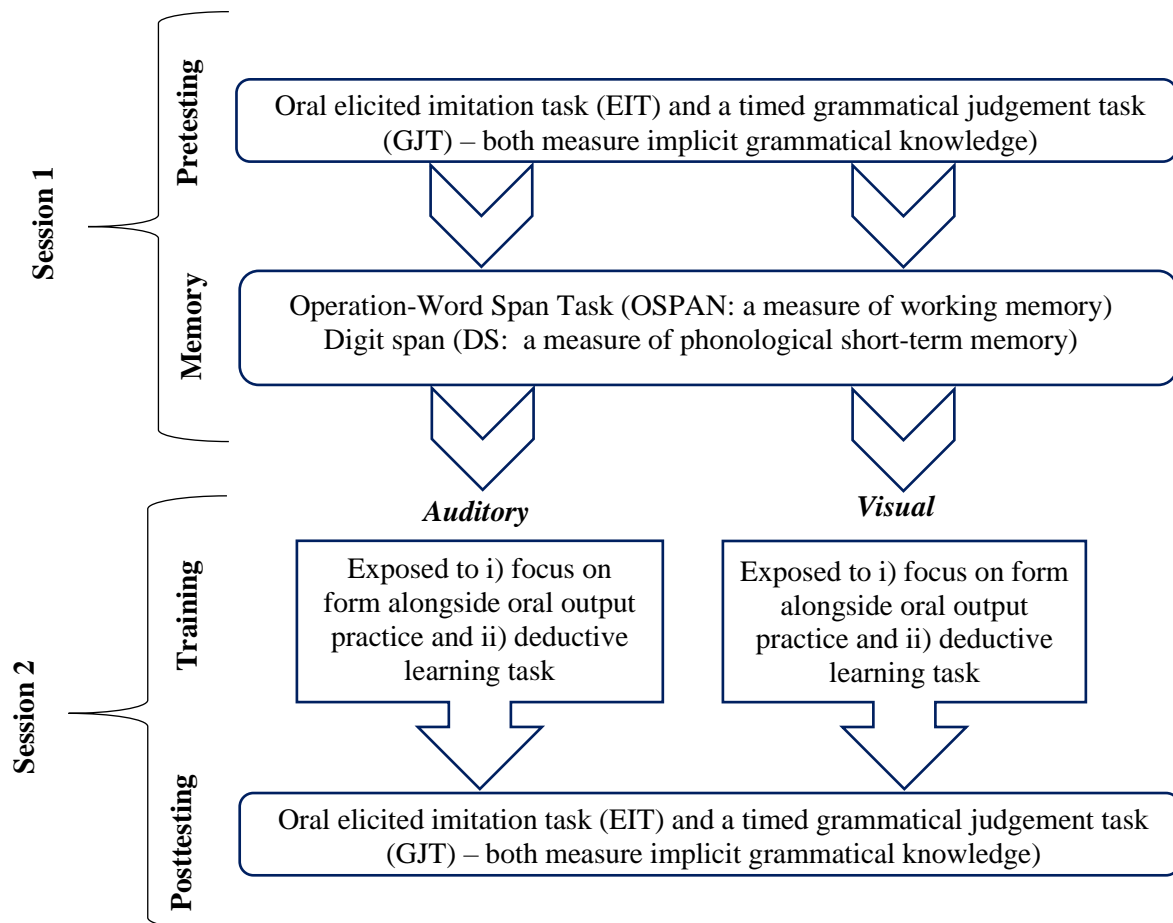
Figure 3. A concise overview of research design.

3.5   Procedure

The experiment was administered to participants mostly individually and, in a few cases, in pairs. A participant, wearing a headset, was seated in front of a laptop computer in a quiet laboratory or room. PsychoPy computer software (Version 1.85.4) was employed to control the stimulus presentation and response collection in the EIT, the timed GJT, and the DS task, while E-prime software (Version 2.0) was used to run OSPAN.

3.5.1 Training stage

Two tasks (i.e., focus on form task and deductive learning task) were employed in the training stage where the input in both tasks was represented auditorily in the auditory group while the same input was represented visually in the visual group. The participants were first exposed to the focus on form task followed by the deductive learning task. Both tasks were performed without time pressure.

3.5.1.1 Focus on form task

The focus on form task involves students' attention being apparently drawn to the syntactic structures as they arise incidentally in an activity where the overriding focus is on semantic meaning. In this research, the participants were exposed (auditorily or visually) to 36 sentences randomly on an item-by-item basis, where 12 sentences targeted each of the three stimuli structures: TQ (n = 12), NA (n = 12), and CON (n = 12). They were aware of neither the underlying goal of the learning task nor the target knowledge that was learned or tested. The frequency of sentences containing the target structure was assumed to induce the participant to notice the target structure (i.e., the learners' attention is attracted by the frequent feature). To ensure that the participants primarily processed a sentence for meaning rather than linguistic form, each statement they were exposed to was followed with two pictures, one of which depicted the sentence. The participants were asked to judge which picture was most compatible with the meaning of the sentence. While listening/reading, the participants did not know which part of the sentence the following pictures described. This approach implies that the participants must process the entire auditory or visual string for meaning. After they had judged the picture plausibility of each statement, they were then required to repeat it orally.

All sentences in this task (and in the deductive learning task) were syntactically correct and semantically plausible. Note: 'semantically plausible' indicates that the meaning

of a sentence is sensible for the world in which we live. However, 'semantically implausible'
means that the meaning of the sentence does not make sense, as in *your laptop minced the
floor yesterday, didn't it?* This sentence is grammatically correct, but semantically
implausible. The task did not commence with practice trials because the task is similar to the
EIT. See Figure 6 as an example of one auditory trial where time pressure is not involved in
the training stage.

3.5.1.2 Deductive learning task

While the focus of the previous task (i.e., focus on form) is primarily on meaning, the
focus of the deductive learning task is primarily on form. Participants in this task were
exposed to 45 novel sentences on an item-by-item basis. While the 36 sentences in the focus
on form task were presented randomly, the sentences in the deductive learning task were
organised into nine groups whereby each group included five trials targeting one of the three
structures. To reduce order effects, the nine groups were placed into a counter-balanced
design generating three blocks, as shown in Figure 4. In the first block (i.e., the first three
groups), a rule explanation of the targeted structure was visually provided while the
remaining groups were presented without the explanation because it had already been
provided.

Block 1 (rule supplied): NA → CON → TQ

Block 2: CON → TQ → NA

Block 3: TQ → NA → CON
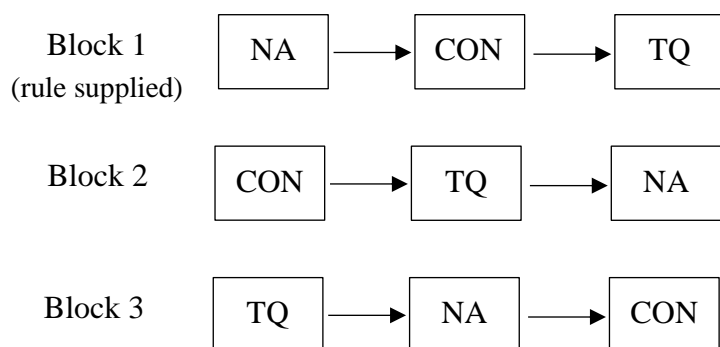
Figure 4. Target structures counterbalanced across three blocks.

The explicit rule explanation provided in block 1 was not detailed so that the demands imposed by retrieval processing of the rule explanation were reduced during doing the trials and as a result, grammatical knowledge of the target structure was promoted to become available for use in automatic processing. Table 2 provides an overview of the rule explanation of each target structure. In each trial, the participants were first exposed (visually or auditory) to a sentence, followed by a simple question about the sentence (e.g., 'Does the sentence you read contain the word '…….'?'), and finally, they were asked to repeat the sentence orally. In this task, the plausibility of the pictures was not included to reduce the distractible effects and increase the amount of attention required to process form. Figure 5 shows an example of one trial from the deductive learning task:

Beginning of trial

- **Stimulus**: *Had he known the answer, he would have told her*

- **Question**: Does the sentence you read contain the word '*answer*'?

Yes   No
Not sure

- **Repetition**

End of trial
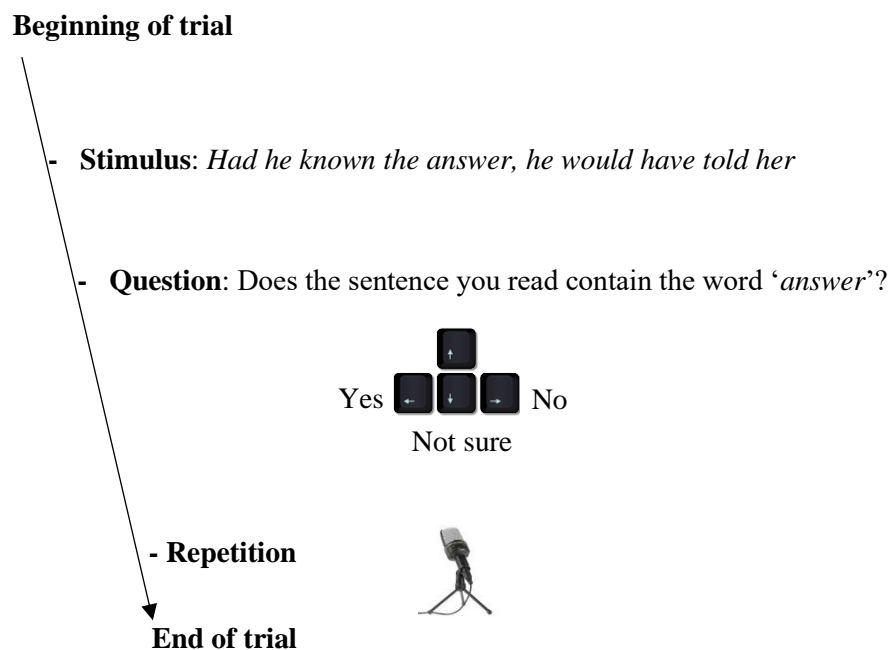
Figure 5. An example of one trial from the deductive learning task.

Table 2

Rule Explanation for Each Syntactic Structure

Structure	Rule explanation
TQ	*Read the sentences carefully and pay attention to where the auxiliary verb comes in each sentence. For example, in the sentence, 'Seldom have they seen such tall buildings', the auxiliary is 'have' and it comes before the subject of the sentence, 'they'.*
NA	*Read the sentences carefully and pay attention to the question added to the end of the sentence, which requires the inversion of the subject and the auxiliary verb (or the verb). For example, in the sentence, 'You have taken my book, haven't you?', you will find that the auxiliary verb, 'have', and the subject, 'you', are inverted in the question.* *Remember that positive sentences are followed by negative questions and negative sentences are followed by positive questions. The sentence, 'You have taken my book', is positive, and therefore the question, 'haven't you?' must be negative.*
CON	*Read the sentences carefully and pay attention to where the auxiliary verb (had) comes in the first part of the sentence and how the grammatical structure in the second part is made (would + have + past participle).* *For example, in the sentence, 'Had I read the instructions carefully, I would have avoided the problem', the auxiliary is 'had' (meaning 'if'), which comes at the beginning of the sentence, and the grammatical structure 'would have avoided' followed the rule provided above (would + have + past participle).*

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

3.5.2   Language testing stage

The participants in each group were pretested/posttested using two measures (EIT and timed GJT) that were designed to measure implicit grammatical knowledge. These two measures are crucial to use together in the current investigation; had only an EIT been employed to measure participants' performance and better performance was obtained from the auditory group, it might not indicate that auditory input was superior. Rather, it would suggest that it was a 'match' case between training and testing. That is, EIT alone could privilege the auditory input modality. Thus, it is necessary to employ aural and visual

measures of implicit knowledge to prevent prejudice in favour of one input modality during learning. In order to reduce learners' use of explicit knowledge on the implicit tests, the EIT was first administered, followed by the timed GJT.

The testing sentences in both measures consisted of novel test items presented to each participant in random order (i.e., the testing items were not employed during the exposure stage). All of them were semantically plausible. The length of statements in both tasks varied between 8 and 18 syllables. This confirms that the statements would include 'stimuli of various lengths', as recommended by Bley-Vroman and Chaudron (1994, p. 253).

In order to reduce the potential practice effects of testing items, two versions of each task (i.e., EIT and timed GJT) were designed where counterbalancing was applied. That is, if a participant was pretested using version A, he/she would be posttested using version B and vice versa. However, it should be noted that there still might be potential practice effects of broader types, including familiarity with the experimental setting (e.g., test purpose), test format, and test content. Substantial efforts were made to make the two versions as close to each other as possible in terms of layout, difficulty, error type, main structure. The following example illustrates how the two versions were designed to be close to each other:

Version A: *You did not hurry to catch the bus yesterday, didn't you?

Version B: *You did not call the police last night, didn't you?

The aforementioned sentence in both versions is formed in simple past tense where both the declarative statement and the tag question were turned into negative. In addition, most of the sentences, not all, in both versions were similar in terms of singular and plural nouns. However, there is a variety in the length of sentences (i.e., number of syllabus) between the two version as shown in the example above. Each version of the EIT and the time GJT contained an equal number of grammatical and ungrammatical sentences. The grammatical sentences followed the identical syntactic forms as the training sentences as in

CON sentence *Had he known the answer, he would have told her*. The ungrammatical

sentences were similar to the grammatical items, with the following exceptions:

1. In the tag questions (TQ), the value in the statement and the tag question was the same:

affirmative-affirmative (AFI/AFI) or negative-negative (NEG/NEG; e.g., *The film finished

much earlier last night, did it?*, or the subject pronoun (SP) in the tag question was swapped

with the auxiliary verb (AUX; e.g., * *The trainer has broken his arm recently, he hasn't?*

2. In the negative adverbs (NA), the AUX was positioned before the NA (e.g., * *Did rarely

Helen clean the house on workdays.*), or the AUX was swapped with the subject (SUB)

(e.g., * *Rarely the students have received feedback on their projects*).

3. In the conditional (CON), the SUB in the second clause was directly followed by the

verb (e.g., *Had you left home early, you avoided the traffic jam.*), or the identical SUB was

followed by 'would' plus the verb (e.g., * *Had John found their address, he would send

their invitation*).

Table 3

An Overview of Ungrammatical Patterns in the Testing Stage; Errors in Bold

Rule	Ungrammatical Pattern	Stimulus
TQ	AFI/AFI or NEG/NEG	*The film finished much earlier last night, **did it**? *The film did not finish much earlier last night, **didn't it**?
	AUX – SP inversion in the tag	*The trainer has broken his arm recently, **he hasn't**?
NA	NA – AUX inversion	***Did rarely** Helen clean the house on workdays.
	AUX – SUB inversion	***Rarely the students have** received comments on their projects.
CON	SUB + V	*Had you left home early, you **avoided** the traffic jam.
	AUX + V	*Had John found their address, he **would send** their invitation.

Note. NA=Negative adverbs. CON= Conditional. TQ = Tag questions. AUX = Auxiliary. SP = Subject pronoun. V = Verb. AFI = Affirmative. NEG = Negative.

Each version of EIT and timed GJT commenced with three practice trials to familiarise participants with the task. The practice sentences were not repeated in the actual testing sets. However, the practice sentences were the same for both versions of each task (EIT and timed GJT).

3.5.2.1 Elicited imitation task

In the EIT, the participants listened to 32 novel sentences that were evenly divided amongst the three target structures and distractors. That is, eight stimulus sentences were created for each of the three target structures. Additionally, eight distractor items were included that consisted of sentences with other various structures relating to relative clause, since/for, third person -s, and auxiliary verb 'be' in the present progressive tense. The items in each structure, including the distractors, were divided into grammatical ($n = 4$) and ungrammatical statements ($n = 4$). To minimise the possibility that the participant monitored the correct regularity governing the stimulus, inducing him/her to respond to the semantic meaning of the utterance, picture plausibility was involved, which is similar to the approach used in the focus on form task in the training stage. However, the procedure involved in the EIT was performed under a time limit. That is, the two pictures were shown for only 6 seconds in order for participants to make a judgement. The participants were instructed in advance to repeat the sentence as quickly as possible (Ellis, 2009; Erlam, 2006). A time limit was established for each utterance based on the length of the utterance. Then, 3 seconds were added to the length of each utterance to allow for the slower processing speed of L2 learners. That is, if the utterance (the sentence the participant was required to listen to) lasted for 4 seconds, the time lag given for repeating the sentence was 7 seconds. The participant's response was recorded. After each trial, there was no time limit, and the participants were instructed to press the 'space bar' to move to the next trial when they were ready. The participants were not informed that the task included ungrammatical statements; however,

they were instructed at the beginning to repeat the sentences in correct English. The underlying assumption in this task is that the spontaneous correction of the ungrammatical forms under time pressure is a strong enough indication that a participant has internalised the target structure.
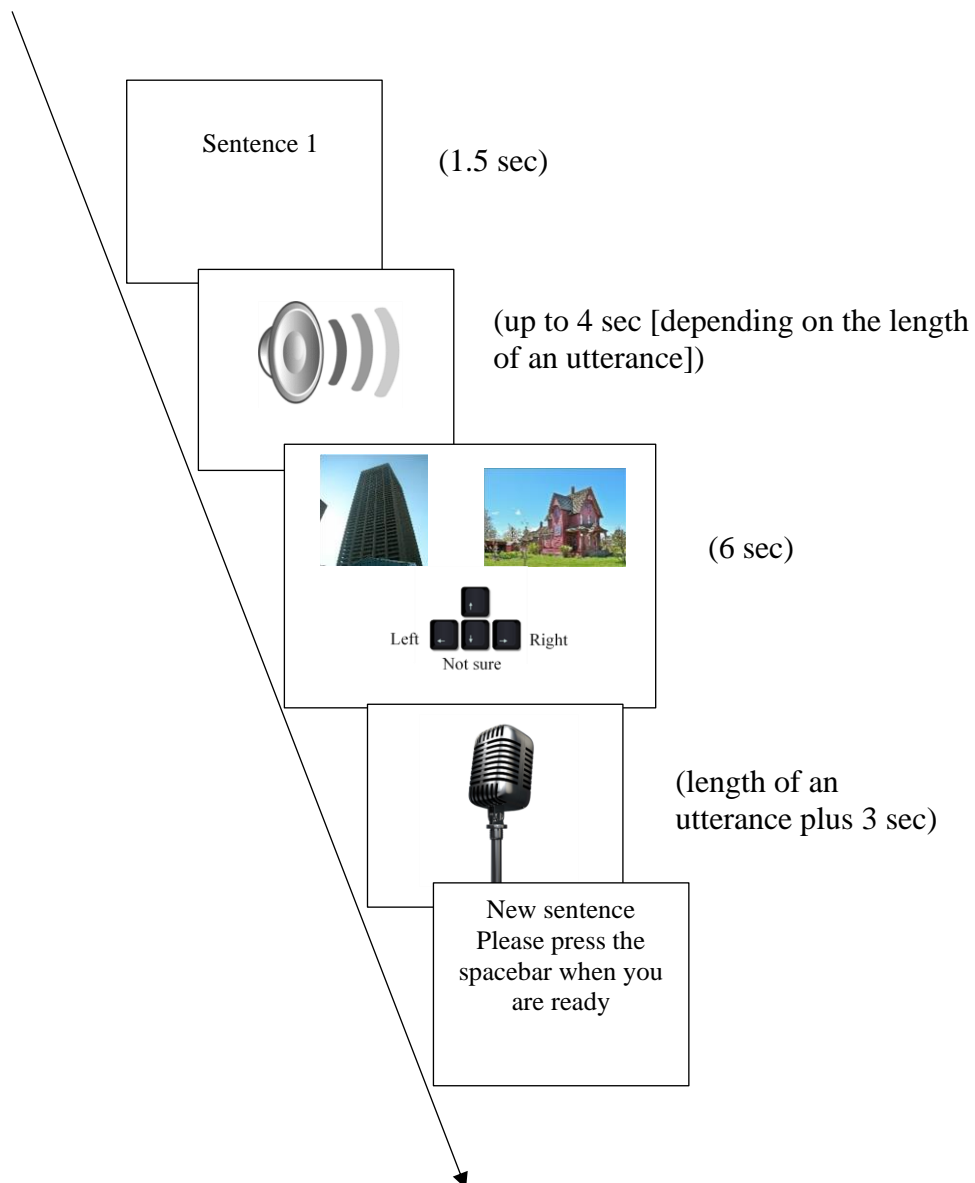
Figure 6. An example of one auditory trial from focus in form task.

Participants' responses were scored based on several types. As scoring the EIT responses is non-trivial, more detailed scoring types with examples are described.

Responses counted as 'correct'. Participants' responses were considered to be correct if they fell into one of the three types below:

Correct-type 1. Grammatical utterances were scored correctly if the participant produced an obligatory occasion using the target structure and it was correctly repeated regardless of lexical and phonological accuracy, as in the example 1 given below:

Example 1

Stimulus: *You remember the <u>English</u> lesson from last week, don't you?*

Response: *You remember the <u>history</u> lesson from last week, don't you?* **(Correct)**

Correct-type 2. Responses to ungrammatical stimuli were scored as correct if the target structure was spontaneously corrected, as in example 2 given below:

Example 2

Stimulus: * *Seldom **wait** a patient **does** in the emergency department.*

Response: *Seldom does a patient wait in the emergency department.* **(Correct)**

Correct-type 3. Responses to ungrammatical stimuli were scored as correct if the target structure was spontaneously corrected regardless if there is a grammatical error in a different position (not in the target structure), as in example 3 below:

Example 3

Stimulus: *These neighbours are moving to the new house, **aren't they**?*

Response: * *These neighbours _moving to the new house, **aren't they**?* **(Correct)**

Correct-type 4. Responses to ungrammatical TQ stimuli were scored as correct if the target structure was spontaneously corrected regardless if the tense is different in stimulus and response.

Example 4

Stimulus: **The cinema was packed with children, was it?*

Response: *The cinema is packed with children, isn't it?* **(Correct)**

Correct-type 5. Responses to ungrammatical NA stimuli were scored as correct if the target structure was spontaneously corrected regardless of the subject-verb agreement. However, the auxiliary verb of the same category (e.g., does, do, and did) should be used and placed in the correct position.

Example 5

Stimulus: **Seldom patients do wait in the emergency department.*

Response: **Seldom does patients wait in the emergency department.*

Responses counted as 'incorrect'. Participants' response was considered to be incorrect if it fell into one of the three types below:

Incorrect-type 1. The first type occurred when a participant produced a response creating an obligatory occasion for the use of the stimulus structure, but either used it incorrectly or pronounced it unclearly such that the stimulus structure could not be understood, as in the example given below:

Example 1

Stimulus: **Seldom wait a patient does in the emergency department.*

Response: *Seldom a patient waits in the emergency department.* **(Incorrect)**

In the example above, the response lacks the auxiliary verb 'does' positioned prior to the subject 'a patient', and the third person -s must be deleted.

Example 2

Stimulus: *You could not lend your iPad to me by any chance, could you?*

Response: **You could not lend your iPad to me by any chance, would you?* **(Incorrect)**

In the example above, as the focus of the ungrammatical pattern in the tag question is that the auxiliary verb must agree with the tense, aspect, and modality of the declarative sentence, the response was scored incorrect because the axillary 'could' was used in place of 'would'.

Incorrect-type 2. - This type occurred when a participant produced a response that did not include an occasion for the obligatory use of the target structure, as in the example given below:

Stimulus: *Seldom <u>does</u> a patient <u>wait</u> in the emergency department.*

Response: *Seldom does a patient* **(Incorrect)**

Supplying an incorrect structure or avoiding the target structure in the second and third criteria was scored as zero.

Incorrect-type 3. This type is linked to the time limit. That is, if a participant could not complete their response, specifically the target structure, within the allocated time limit, the current trial would be counted as 'incorrect' (i.e. any oral response beyond the time limit is simply not recorded).

Cronbach's alpha was used to calculate reliability for the 24 items (excluding filler sentences) for EIT's version A and B in both experiments. Scores were computed several times as shown in Table 4 and 5.

Table 4

Reliability Alpha Scores for the Version A and B of EIT Scores for Experiment 1

Reliability test	Pretest		Posttest	
	A	B	A	B
α	0.68	0.64	0.79	0.78

Table 5

Reliability Alpha Scores for the Version A and B of EIT for

Experiment 2

Reliability test	Pretest		Posttest	
	A	B	A	B
α	0.67	0.57	0.71	0.81

3.5.2.2 Grammatical judgement task

The present investigation also involved the use of a timed GJT, which is a timed test and is always presented in the visual mode. The task consisted of 32 novel sentences so that each to the three target structures was presented eight times including eight distractor sentences. Half of the stimulus items for each structure were grammatical ($n = 4$), and the other half were ungrammatical ($n = 4$). In order to establish the time limit necessary to judge each sentence, the GJT was trialled on 10 native speakers of English by timing their performance on each item. They were instructed to respond as quickly as possible, and their response times were recorded. An additional 20% of the time was added to each allow for the slower processing speed of L2 intermediate-level participants (R. Ellis, 2005; Loewen, 2009).

The participants were required to read each sentence and judge its grammaticality by choosing one of three possible responses: 'correct', 'incorrect' or 'not-sure'. The third option, 'not sure', was added because if a participant guessed the option correctly more by luck than by judgement, then he or she could obtain a high score. This threat, statistically, implied that the scores in the distribution would regress to the mean as a result of guesswork and not of certainty and automatized knowledge (Gravetter & Forzano, 2015). Upon responding to the current sentence, the participant automatically moved on to the next sentence. Similarly, as described in the EIT, the participants were instructed to respond as quickly and accurately as possible based on their first impression. Responses were scored dichotomously as correct (one point) or incorrect (zero points), with items to which the 'not sure' response or no

response was given being treated as incorrect. If a participant could not judge the stimulus within the allocated time limit, the current trial would be counted as 'incorrect' and the next trial would begin automatically. The participants were given one self-paced break.

As noted earlier, the pretest performance in the timed GJT was the main measure which was used to determine the extent of the participants' prior knowledge of the stimulus material. Accordingly, two participants were eliminated from the analysis as they obtained above 19 out of 24 in total, where a score above 19 was assumed to indicate that a participant has good grammatical implicit knowledge. One additional participant was also removed from the analysis because the participant performed the posttests twice by mistake.

Cronbach's alpha was used to calculate the reliability for the 24 items (excluding filler sentences) for GJT's version A and B in both experiments. Scores were computed several times as shown in Table 6 and 7.

Table 6

Reliability Alpha Scores for the Version A and B of EIT for Experiment 1

Reliability test	Pretest		Posttest	
	A	B	A	B
α	0.58	0.44	0.51	0.71

Table 7

Reliability Alpha Scores for the Version A and B of EIT for Experiment 2

Reliability test	Pretest		Posttest	
	A	B	A	B
α	0.67	0.57	0.71	0.81

3.5.3 Memory testing stage

The participants in both experimental groups performed an OSPAN task as a measure of WM capacity, followed by a DS test as a measure of phonological short-term memory.

3.5.3.1 Operation span test

The OSPAN test is based on a version created by Unsworth et al. (2005). It is a complex task which is designed to measure WM capacity, as it involves not only storage but also a degree of information processing. That is, the test requires participants to read and verify a simple mathematical problem, such as 'Is $(4/2) - 1 = 1$?', and then read a letter after the operation (such as *M*). They are presented with a total of 15 problem sets where the size of each set ranged from 3 to 7 letters and mathematical problems, generating a total of 75 letters and 75 mathematical problems. At the end of each set, participants were required to remember the letters that followed each mathematical problem in the order presented.

The participants were awarded one point for every letter remembered in the correct order. If a participant took longer than the average amount of time to solve a mathematical problem, the program would automatically move him/her on to the word part or the next mathematical problem, and then the mathematical problem would be counted as a mathematical error. A number was shown at the end of each set to indicate the percentage of correct answers a participant had obtained for the mathematical problems. Only participants who had achieved a score of at least 80 % on the mathematical problems were included in the analysis that involved the WM measures. This ensured that both storage and processing components of memory were occupied during the assessment (Tagarelli et al., 2015). The entire task lasted approximately 35 minutes.

3.5.3.2 Digit span (DS) test

The DS was designed based on a version created by Gathercole, Service, Hitch, Adams, & Martin (1999). The DS is a simple task, which is designed to measure phonological short-term memory. Unlike OSPAN, the DS measures immediate recall for a forward sequence of digits (i.e., only the storage component is involved). As mentioned earlier, the DS in the present investigation was used only for the purpose of controlling the effects of short-term memory capacity on EIT performance. In a DS task, the participant hears a sequence of digits and is asked to recall them immediately in the identical forward sequence. The digits range from one to nine and are presented in a random order. Two lists of digits are included at each trial, starting with two digits. The DS is scored based on a number of determinants. If two lists in each trial are correctly recalled, then the length of the next list is increased by one, and the other two lists will be added. If the subject fails to recall both of the two lists at one trial, then no further lists are given. However, if a participant succeeds in remembering either of the two lists, then a third list is added at that trial. If the third one is correctly recalled, then a new trial is shown. If the subject incorrectly recalls the third list, then the test will automatically stop.

Chapter 4: Experiment 1

4.1   Introduction

The aims of this experiment were twofold. The first was to determine whether there is a significant difference in the relative effects of auditory and visual input modalities on the acquisition of implicit grammatical knowledge. The second aim was to determine whether WM capacity mediates the effects of auditory or visual input modality on the acquisition of implicit grammatical knowledge. To reduce the possibility of considerable variations in the results due to participants' L1, participants in Experiment 1 were limited to Chinese speakers while participants in Experiment 2 were limited to Arabic speakers. Therefore, the theoretical basis in relation to Chinese syntax will be addressed in this chapter, focusing on the Chinese counterparts of English stimuli structures used in the current investigation, followed by a comparison between English and Chinese syntax in terms of the stimuli structures. It must be pointed out, however, that because L1 interference is not the focus of the current study, and the tested structures were not selected due to their different representations in the participants' L1 (Chinese vs. Arabic), the comparison of the target syntax between the two languages will not be examined in detail, but such background information about the two languages is informative for readers.

4.2   Participants

A total of 87 participants, that is, ten native speakers of English and 77 non-native speakers of English, took part in this study. The native speakers of English were postgraduate students at the university, who were employed only for establishing the time limit required in the timed GJT. The non-native speakers were Chinese native speaker participants, who were randomly assigned to one of two groups: the auditory group (7 males and 31 females, aged 20–30 years '$M = 23$') and the visual group (7 males and 32 females, aged 20–30 years '$M =$

23'). Most of the non-native speaker participants were students who were either studying English in Bristol or who were a part of an undergraduate or postgraduate degree programme. They were intermediate-level speakers of English. This English level is the equivalent of a score of 5.5–6.5 on the academic version of the International English Testing System (IELTS) or equivalent scores on the Test of English as a Foreign Language (TOEFL, no participants had a TOFEL score in Experiment 1), provided that it was the last score that a participant received. The period between the last IELTS score and the time when the participant took part in the experiment was not strictly managed in order to maximise the opportunity of recruiting as many potential participants as possible. However, the IELTS score had to have been gained within the last 2 years at the most. More importantly, the pretest performance on the timed GJT was the main measure which was used to determine the extent of the participants' prior knowledge of the stimulus material (see Section 3.5.2.2, for information about the elimination of some participants). Participants received £14 for participating in the experiment.

4.3   Chinese syntax

The Chinese language belongs to the Sino-Tibetan family of languages. It is one of the largest language families in the world; its number of first-language speakers is more than Indo-European (Pereltsvaig, 2017). Linguists use the term Sinitic to describe Chinese languages, including dialects which are distinct from one another yet similar, analogous to the similarities and differences between European languages such as Romanian and Portuguese (Chappell & Lan, 2016). Around one billion people, approximately one-fifth of the Earth's population, speak some form of Chinese (for a review, see Chappell & Lan, 2016). Chinese consists of a number of dialects, some of which are so different that they cannot be mutually understandable, and linguists consider them distinct languages. These

dialects can be sorted into a minimum of five groups according to their structural affinities (see Thompson & Li, 2009, for a review about the five groups):

Mandarin (an English translation of the old Beijing expression *guan-huà* [official language]) is the main dialect spoken in China. Its native speakers constitute about 70% of China's population (Joseph, 2019). Mandarin is the major dialect spoken by people in the northern, central, and southwestern parts of the China (Joseph, 2019).

Other vital dialects include Wú and Mˇin. The former is spoken in Shanghai, the largest city in China, while the latter is spoken in Táiwa-n and Fújiàn provinces and Haˇinán Island in the Gulf of Tonkin. The fourth group is the Yuè dialects which are spoken mainly in the area of Guaˇngdo-ng. Yuè dialects, including Cantonese, the language of Guaˇngzho-u (Canton), are also spoken in several areas of the Chinese diaspora, specifically Hong Kong and overseas Chinese settlements such as the Chinatowns in the United States, Europe, and Southeast Asia. Therefore, many English words which are borrowed from Chinese have their origins in Cantonese, for example, kumquat from the Cantonese *kamkwat*.

The final dialect is Hakka which is well known outside of China due to the emigration of its speakers from China. Most Hakka people are dispersed across southeastern China in Guaˇngxı- province and across the Mˇın and Yuè areas.

Although the five dialect groups have considerable differences with regard to both vocabulary and grammar, Chinese people basically share a uniform written language which is logographic. Put differently, Cantonese and Hakka speakers cannot contact each other orally either over the phone or in a conversation face to face, but they can understand each other in writing and can both comprehend newspapers (Joseph, 2019). The following is an example of the difference between a Chinese sentence as spoken in Hakka and Cantonese:

Table 8

An Illustration of Spoken Hakka and Cantonese Dialects and their Shared Written Form

	Spoken Hakka	Spoken Cantonese	Written Chinese
Sentence	崖食等饭	我 而家 食 緊 飯	我 现 在 吃 饭
Transliteration	Ngai si den fan	Ngo yiga sek gang fan	Wo xianzai zai chi fan
Gloss	I/eat/ -ing/rice	I/now/eat/am -ing/rice	I/now/am -ing/eat/rice
Translation	I am eating rice now.	I am eating rice now.	I am eating rice now.

Note. The example sentence of spoken Cantonese and written Chinese was taken from Arc communications (2011).

Table 8 illustrates large pronunciation differences between the Hakka and Cantonese dialects. However, both share the same written language which is in the form of standard Chinese (Mandarin), irrespective of whether simplified or traditional Chinese characters are applied.

Chinese written language is based on characters rather than letters. These characters, which have progressed over time from symbolic pictures, portray meaning more than sound. As a result, the Chinese people who speak different dialects often pronounce the identical written character very distinctly. Unlike English and Arabic, both of which have letters that encode an ortho-phonological regularity, Chinese does not have an alphabet in which both the meaning and pronunciation of characters can be learned only by memorization. Chinese includes approximately 50,000 distinct characters, but there are around 4,000 basic characters which a Chinese speaker must possess to read and write. This is due to the fact that most of these characters are archaic and have become less common over time. Therefore, Chinese must be transliterated if native speakers of English, for instance, wish to become familiar with how the Chinese characters are pronounced (Joseph, 2019).

While many languages, including Arabic, English, French, and Russian, normally have inflectional morphemes to assign a word specific grammatical features such as the

tense/aspect and number/person of the subject, definiteness, indefiniteness, or comparison, Chinese dialects do not have such inflectional categories. Most Chinese words are made of one or two morphemes. More specifically, Chinese language includes few inflectional morphemes. The type of morphology inherent in Chinese dialects tends to require compound (two or more free morphemes joined together) and derivational morphemes (which often change the part of speech) rather than inflectional ones. This type of morphology is particularly dominant in the modern Mandarin dialect (Li & Thompson, 2009). For example, the verb *lai* [to come] remains unchanged even if the action occurred in the past, is happening currently, or will happen in the future. Furthermore, the verb remains the same in form regardless of whether the subject is the third person singular or the f3 and 4irst-person plural. This means, overall, a verb in Chinese is consistent in form. Another good example can be reflected in the verb *to be* in that it has many different forms in English based on the tense or person (e.g., am, is, are, was, were, will, etc.), while Chinese has just one counterpart, *shi* (Jiang, 2009).

The explicit lack of inflectional morphemes to indicate tense in Chinese verbs, agent-patient roles in a pronoun, and subjunctive mood is remedied by word order variation in the sentence's information structure (Kirkpatrick, 1993, as cited in Jiang, 2009). As a result, word order variations in Chinese sentences have more intricate grammatical roles than those in English. Consequently, word order governed by syntax in Chinese is one of the most striking aspects that distinguishes it from English. Ho (1993, as cited in Jinag, 2009) asserted that Chinese is "one of those languages that rely heavily on word order as an underlying marking feature for meaning". Word order variation in Chinese not only plays a role in altering meaning, but it also reveals the brilliant control of the words in a sentence. The *shunkouliur* [Chinese doggerel] in the following example demonstrates that the citizens of the

regions vary according to their tolerance of spicy food (note: examples were taken from Jiang, 2009).

1. 四川 人 不 怕 辣. (simplified language)

    Sichuan ren **bu pa la**.

    Sichuan person **not fear spicy**.

2. 湖北 人 辣 不 怕

    Hubei ren **la bu pa**

    Hubei person **spicy not fear.**

3. 湖南 人 怕 不 辣

    Hunan ren **pa bu la**

    Hunan person **fear not spicy.**

The variation in word order, specifically in the last three words of each sentence, suggests that people from the Hunan region (sentence 3) bear the spiciest food; the Hubei people come next (sentence 2), followed by the Sichuan people (sentence 1). This delicate distinction in meaning results only from the rearrangement of the word order of the last three syllables in each sentence: *bu* [not], *pa* [fear], and *la* [spicy food]. A closer look at the three characters indicates the differences in meaning pertain to the position of negation, the verb, and the noun. The first sentence's word order (bu/pa/la 'negation + verb + noun') is expressed 'people do not fear something'. The second sentence's word order (la/bu/pa 'noun + negation + verb') connotes the noun is emphasised, which suggests that it is the least people fear. In the third sentence's word order (pa/bu/la 'verb + negation + noun'), the negation does not refer to the verb; instead, it negates the noun. This word order suggests that people love food with spice and that they 'fear' their food not being spicy. As noted, the distinctions in meaning cannot be explicit when translated into English (for further review, see Jiang, 2009).

According to Greenberg (1963, as cited in Jiang, 2009, p.7), word order in world languages can be classified on the basis of the placement of the three parts of a declarative transitive sentence: subject (S), verb (V), and object (O). Tomlin (2014) suggested the comparative frequencies of six canonical word orders of the three constituents in languages: SOV, SVO, VSO, VOS, OVS, and OSV, where the most common word orders are SOV and SVO. The English language uses a firm SVO word order. Thompson (1978, as cited in Jiang, 2014, p.7) explained:

> English is a language in which basic grammatical relations are signalled by word order. Specifically, it is a language in which there must be a noun phrase [NP] immediately preceding the verb in main clauses and that noun phrase, if unmarked, is the subject.

> To verify this description, sentences with a 'dummy' pronoun subject such as *it* or *there* substantiate English's inflexible SVO word order. For instance, in declarative clauses like 'it is interesting to see you again', the 'dummy' subject *it* is prerequisite to occupy the subject position although it does not bear any lexical meaning.

The Chinese language uses SVO or SOV word order although there is fierce debate over which word order is dominant. Tai (1973, as cited in Jiang, 2009) claimed that the canonical word order of Chinese is SOV on account of the idiosyncrasies inherent in SOV languages as determined by Greenberg (1963, as cited in Jiang, 2009). These idiosyncrasies include the sequence of modifiers prior to the modified, the frequent use of postpositions, and the application of interrogatory helping words at the end of Yes–No questions (Ho, 1993 as cited in Jiang, 2009). Because all these features exist in Chinese, Tai deduced that Chinese is an SOV word order language. Others linguists such as McGinnis (1988) and Lu (1980) proposed that Chinese is chiefly an SVO language on the basis of a number of statistical studies carried out scrutinising the frequency of SVO clauses (e.g., Sun & Givón, 1985; M.

Wang, 1988, see Jiang, 2009, for a review). While a great number of researchers hold the position that Chinese is mainly an SVO language with regard to statistical predominance and unmarked surface level word order, the word order that best describes Chinese is still unclear because it carries both SVO and SOV features (for comprehensive discussion about word order, see Jiang, 2009).

4.3.1   Relevant syntactic structures in Chinese

4.3.1.1 Tag question

Whilst English has positive and negative tags specified based on whether the statement is negative or positive, Chinese can freely possess three tag patterns—verb-not-verb, verb-particle, and negative-verb-particle—regardless of whether the preceding statement is positive or negative. Each pattern functions for distinct discourse purposes.

Unlike English tags, in which the tag verbs are formed with an auxiliary or modal verb derived from the statement, Chinese possesses a limited number of tag verbs, such as *shi*, *dui*, *hao*, *xing*, *keyi*, *you*, etc. The tag verbs are basically discourse verbs for replying to a statement in which they normally serve various discourse functions (Hsin, 2016, see for a comprehensive review). In sentence 1 below, for instance, *shi* suggests that Speaker B agrees with the information of Speaker A's proposition. However, *dui* in sentence 2 shows that Speaker B confirms the value of Speaker A's statement.

1. A: 语言学很难学吗？B：是，很难

   A: yuyan-xue hen-nan xue ma? B: shi, hen-nan.

   A: Is linguistics very hard to learn? B: Yes, it is hard.

2. A: 明天要开会吗？ B：对呀，十点。

   A: mingtian yao kaihui ma? B: dui-ya, shi-dian.

A: Do we have a meeting tomorrow? B: Yes. Ten o'clock.

Finally, unlike English tags, which involve a pronoun subject, Chinese tag questions do not require a subject and entail only a verb, as illustrated in the Chinese tag examples in sentences 3 and 4.

3. 张三已经走了 （是不是？/是吗？/不是吗？）(positive statement)

 Zhangsan yijing zou-le, (shi-bu-shi? / shi-ma? / bu-shi-ma?)

 Zhangsan has gone, (hasn't he? / is it? / isn't it?)

4. 张三没来，（是不是？/是吗？/ 不是吗？ (negative statement)

 Zhangsan mei-lai, (shi-bu-shi? / shi-ma? / bu-shi-ma?)

Zhangsan did not come, (did he? / is it? / isn't it?)

(note: the examples above were taken from Hsin, 2016)

4.3.1.2 Conditionals

Unlike English conditionals (CON), in which the degree of hypotheticality, such as factual, hypothetical, or counterfactual, is determined by distinct forms of tense and modality in the verb phrase, Chinese conditionals have only one verb-tense form to express different degrees of hypotheticality. That is, Chinese speakers do not need to alter tense and modality in the verb phrase to demonstrate whether the conditional sentence is factual, hypothetical, or counterfactual. Instead, Chinese makes use of the context, time reference (such as 'last week' or 'now'), and internal semantic logic to differentiate the conditional types (Wu, 1994, as cited in Chou, 2000). According to Li & Thompson (1989, p. 647), the hearer deduces the exact type of conditional message "from the proposition in the second clause, and from his/her knowledge of the world, and of the context in which the sentence is being used". Counterfactual conditionals in Chinese, more specifically, have four basic factors through which counterfactual sentences are expressed: temporal references, aspect markers, negators,

and rhetorical interrogatives (Chen, 1988; Wu, 1994, as both cited in Chou, 2000). Temporal references, which are the most crucial element in Chinese counterfactual conditionals, are lexical constituents that indicate time, such as those having meanings such as 'last week' or 'currently'. Examples 1 and 2 below show how a single temporal reference word, *zuotian* [yesterday], changes the sentence from a future predictive to a past counterfactual conditional (Note: all example sentences below were adapted from Wu's (1994) book, as cited in Chou, 2000).

1. Yaoshi ni bang ta, ta hui qu de. (clause-final particle).

    If you help him, he will go.

2. Zuotian, yaoshi ni bang ta, ta hui qu de.

    Yesterday, if you had helped him, he would have gone.

    The typical Chinese aspect marker for counterfactuality is the verbal particle *le*, which indicates the action is completed as shown in sentence 3 below. When the particle *le* is used together with the time reference word *zao* [earlier than, a long time ago] in the main clause of a conditional, the combination indicates a counterfactual conditional.

3. Huo yaoshi dao le, wo zao jiu tongzhi ni le.

    If the goods had arrived, I would have notified you long ago.

    The third principal factor used to express a counterfactual message is the negator *bu* [not], which modifies the if-words *yaoshi* or *ruguo* or the verb in the if-clause. When the negator *bu* is combined with the time reference word *zuotian* [yesterday] as in 4 below, a counterfactual conditional is expressed.

4. Zuotian, yaoshi/ruguo ni bu bang ta, ta jiu fangqi le.

    Yesterday, if you had not helped him, he would have given up.

The last factor related to counterfactuality is the rhetorical interrogatives, which are unique to the Chinese language. They arise in the pattern of rhetorical questions in the main clause as shown in 5, and the conditional clause is usually negative (Wu, 1994, p. 182).

5. Yaobushi wo, ta hui you jintian?

    If it were not for me, could/would he be what he is now?

However, a conditional sentence in Mandarin – the official and written language – can be made up in the absence of any grammatical marking so it is ambiguous to non-speakers of Chinese to distinguish between the three types of conditionals (i.e., unreal, hypothetical, and counterfactual), as revealed in sentence 6 below.

6. Conditional clause: ruguo ni kan dao   wo   meimei,

 Gloss: If you see arrive my younger sister,

 Main clause: ni   yiding   zhidao ta huaiyun le.

 Gloss: You certainly know her pregnant.

This sentence can be translated as real conditional: *If you see my younger sister, you'll certainly know that she is pregnant*, as hypothetical: *If you saw my younger sister, you would know she was pregnant*, or as Counterfactual: *If you had seen my younger sister, you would have known that she was pregnant.*

Given that evidence, Wu (1994, as cited in Chou, 2000) argued that English and Mandarin Chinese use different levels of linguistic devices to express counterfactuality. English counterfactuals, according to Chou (2000), depend on the linguistic features at the syntactic level, such as subjunctive mood verbs. Chinese, on the other hand, relies on the lexical, syntactic, and discourse levels. The temporal reference words are on the lexical level; the negation and the rhetorical interrogatives are on the syntactic levels.

4.3.1.3 Negative adverbs

Similar to English, Chinese language has negative adverbs (NA), but these adverbs are limited compared to those in English, which possesses a variety of lexical words expressing negative adverbs such as *nowhere*, *never*, *seldom*, *rarely*, *barely*, *scarcely*, *little*, *hardly*, *no sooner . . . than*, *not only. . . but also*, *only when*, and so on. However, negative adverbs in Chinese neither occur at the beginning of the sentence nor provoke the subject-verb inversion in any situation.

4.3.2   L1 Transfer

Considering the comparisons between English and Chinese language in terms of syntax leads to the concept of L1 transfer underlying the contrastive analysis hypothesis which was sparked by Lado in the 1950s. According to Lado (1957, as cited in Gass & Selinker, 2001), language transfer refers to the influence that the learner's L1 exerts over the learning of a second language. His hypothesis suggests that comparing the differences and similarities between the L1 and L2 could predict linguistic areas which are difficult and easy to learn. Lado (1957, as cited in Gass & Selinker, 2001, p.72) argued,

> Those structures that are similar will be easy to learn because they will be transferred and may function satisfactorily in the foreign language. Those structures that are different will be difficult because when transferred they will not function satisfactorily in the foreign language and will therefore have to be changed.

 Similarities between the L1 and L2 refers to positive transfer, while the differences between them refer to negative transfer. In the case of an English–Chinese comparison, the basic word order of English (SVO) behaves in a similar way to Chinese as the word order of the latter can be SVO or SOV. This means that Chinese learners of English likely will not find it difficult to learn English word order.

In terms of stimuli structures used in the current investigation, particularly, the comparisons revealed that the English tag question (TQ) does not behave similarly to the Chinese counterpart as the tag verb in the latter is neither formed with an auxiliary or modal verb derived from the statement nor does it require a pronoun subject as in English. Therefore, the high level of difference in TQ structure between the two languages suggests that the TQ is likely to be difficult to learn for Chinese learners of English. In the case of conditionals (CON), the comparisons showed a high level of difference because the degree of hypotheticality (i.e., factual, hypothetical, and counterfactual) in the Chinese language is determined by the context, time reference (such as 'last week' or 'now'), and internal semantic logic using one verb-tense form. Thus, Chinese learners of English are more likely to find difficulty in learning the CON structure. For the negative adverb (NA) structure, although Chinese has a number of lexical words expressing negative adverbs, these adverbs do not cause subject-verb inversion when the adverb occurs at the beginning of a sentence as they do in English. As the aspect tested is the L2 learners' ability to do so, such structure is likely to be difficult to learn. To put all stimuli structures together, all generated structures used in the current investigation appear likely to be difficult to learn for Chinese learners of English due to L1 interference (i.e., negative transfer).

Table 9

Summarising the Level of Difficulty Based on a Chinese–English Comparison for the Three Structures

Rule	Level of difficulty
NA	High
TQ	High
CON	High

Note. Due to the L1 interference is not the focus of the current investigation, only basic comparisons were aimed at; TQ = tag questions; NA = negative adverb; CON = conditional.

4.4   Results

This section provides a detailed account of the statistical analyses carried out on the data obtained from the experiment conducted to test the hypotheses put forward in the preceding chapter. The first aim of analysis was to investigate whether there is a significant difference in the EIT/timed GJT scores as a function of exposure to auditory input vs. visual input during language learning. This type of analysis allows for demonstrating which of two input modalities (auditory or visual) would elicit more successful acquisition of implicit grammatical knowledge (Research Question 1). The second aim of analysis was to explore whether there is a significant correlation between OSPAN task scores and the increase in performance (from pre-to posttests) on the EIT / timed GJT within each modality group. This type of analysis allows for determining whether WM capacity plays a differential role on success in the acquisition of implicit grammatical knowledge under exposure to the auditory and visual modalities (Research Question 2).

To investigate causality and differences in the data pertaining to the first research question, repeated-measures ANOVAs and mixed-ANOVAs were performed. These analysis allow for a tracking increase in performance from pre-to posttests, combining three variables

in one analysis, such as, in the case of a mixed ANOVA, a within-group variable (testing session: pre-vs posttests), a within-group variable (syntactic structure: TQ vs NA vs CON) and a between-group variable (modality intervention: auditory vs. visual). However, it should be noted that assumption of normality was violated in some groups because data in all groups and sub-groups do not commonly meet the criteria of normality in one analysis. This assumption was commonly violated in previous L2 research (Nasasaji, 2012). However, this violation can be corrected with the non-parametric tests for a t-test and a one-way ANOVA where the increase between pre-and posttests is used as a dependent variable. However, this approach will result in a large series of tests, leading to the risk of an inflated Type 1 error. That is, "the more analysis you run the more likely you are to find a significant result even if in reality there are no differences between your groups" (Pallant, 2016, p. 289). Therefore, a mixed-ANOVA was considered the most appropriate inferential test to examine causality and differences in the data pertaining to the first research question.

The inferential analysis for GJT and EIT data was complemented with the following three goals (in order):

1. To examine how the mean improvements in performance between the pre- and posttests differed between the two interventions (auditory and visual), and whether any difference between the two, if present, was influenced by type of syntactic structure or grammaticality. To achieve this goal, a mixed ANOVA was performed twice: a 2 (testing session: pretest vs posttest) $\times$ 3 (syntactic structure: TQ vs NA vs CON) $\times$ 2 (modality intervention: auditory vs visual) mixed-ANOVA and a 2 (testing session: pretest vs posttest) $\times$ 2 (grammaticality: grammatical vs ungrammatical) $\times$ 2 (modality intervention: auditory vs visual) mixed-ANOVA.

2. To investigate whether auditory intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality. To accomplish this goal, a 2 (testing session: pretest vs posttest) $\times$ 3 (syntactic structure: TQ vs NA vs CON) $\times$ 2 (grammaticality: grammatical vs ungrammatical) repeated-measures ANOVA was performed.

3. To investigate whether visual intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality. To achieve this goal, the same repeated-measures ANOVA used in 'goal 2' was conducted.

To explore correlations in the data pertaining to the second research question, first, simple linear regression was conducted to determine only whether the improvement in performance in GJT/EIT is correlated to the performance in the OSPAN task within each modality group without taking other factors into consideration. Then, multiple linear regressions were performed to find out to which extent the strength of correlation was affected when IELTS scores (in the case of GJT and EIT) and phonological short-term memory (in the case of EIT) were controlled.

4.4.1 Analysis of timed GJT results

Table 10 shows participants' mean raw scores, with standard errors in both the pre- and posttests in each of the two modality-specific interventions, including overall performance and means at each level of the syntactic structure and grammaticality conditions.

Table 10

Correct Mean (M) Raw GJT Scores with Standard Errors (SEs)

Condition	Pretest		Posttest		Improvement	
	M	*SE*	*M*	*SE*	*M*	*SE*
Overall performance	12.49	0.37	16.57	0.38	4.07	0.45
Auditory	12.79	0.45	15.97	0.54	3.18	0.61
Visual	12.21	0.58	17.15	0.53	4.94	0.64
Syntactic structure						
TQ	4.74	0.19	5.99	0.14	1.24	0.22
Auditory	4.92	0.25	5.79	0.20	0.86	0.29
Visual	4.56	0.28	6.18	0.20	1.61	0.31
NA	3.82	0.18	5.01	0.19	1.19	0.21
Auditory	3.71	0.27	4.76	0.27	1.05	0.33
Visual	3.92	0.23	5.26	0.26	1.33	0.27
CON	3.94	0.16	5.57	0.18	1.63	0.22
Auditory	4.16	0.20	5.42	0.27	1.26	0.30
Visual	3.72	0.24	5.72	0.23	2.00	0.31
Grammaticality						
GR	8.08	0.30	10.19	0.23	2.11	0.36
Auditory	8.32	0.42	9.63	0.38	1.31	0.53
Visual	7.85	0.45	10.74	0.26	2.89	0.46
UNG	4.42	0.24	6.38	0.28	1.96	0.31
Auditory	4.47	0.33	6.34	0.41	1.86	0.42
Visual	4.36	0.36	6.41	0.39	2.05	0.46

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items. The maximum raw score is 24 for overall performance, 8 for each syntactic structure, and 12 for each type of grammaticality (i.e., each type of syntactic structure includes four grammatical items and four ungrammatical items, for a total of 24).

Before performing inferential analyses to address the research questions, A Mann-Whitney test was performed on the pretest scores to rule out differences between the auditory and visual groups prior to the intervention. The analysis demonstrated no significant difference in GJT scores for auditory modality (Mdn = 12.5) and visual modality (Mdn = 13.0) conditions, U = 706.00, p = 0.73, indicating that the sample was homogeneous and that any differences observed in the posttests could be attributable to the modality intervention. In addition, it is also important to check the data of each variable used in the ANOVAs below in terms of the normality of distribution and outliers. At first, the pre- and posttest performances for each group were checked individually. The results revealed that the data of the variables that pertained to answering the first research question did not contain any outliers. The results of the Shapiro-Wilks test for normality revealed that the posttest data for each modality group were normally distributed, whereas the pretest data for each group were not. Furthermore, it has been noted that when the data were reduced in size (e.g., pretest performance on grammatical items), the possibility of outliers and abnormal distribution of data appearing was higher. The following tables: Table 11, Table 12, Table 13, Table 14, and Table 15 shows a summary of the data of all variables in terms of normality of distribution and outliers.

Table 11

Summary of Shapiro-Wilk Tests for GJT Data Across Both Modality Groups and in Each
Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Overall performance				
Pretest	0	77	0.97	0.09
Posttest	0	77	0.98	0.25
Modality				
Auditory				
Pretest	0	38	0.93	0.03
Posttest	0	38	0.97	0.39
Visual				
Pretest	0	39	0.93	0.02
Posttest	0	39	0.95	0.12

Table 12

Summary of Shapiro-Wilk Tests for GJT Data on Items of Syntactic Structures Across Groups and Each Individual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Syntactic structure				
TQ				
Pretest	2	77	0.93	0.001
Auditory	1	38	0.93	0.03
Visual	0	39	0.19	0.006
Posttest	0	77	7.89	< 0.001
Auditory	0	38	0.90	0.003
Visual	4	39	0.88	0.001
NA				
Pretest	0	77	0.95	0.009
Auditory	0	38	0.94	0.08
Visual	0	39	0.94	0.06
Posttest	0	77	0.95	0.01
Auditory	0	38	0.95	0.18
Visual	0	39	0.92	0.09
CON				
Pretest	0	77	0.92	< 0.001
Auditory	0	38	0.90	0.003
Visual	0	39	0.92	0.01
Posttest	1	77	0.94	0.002
Auditory	0	38	0.93	0.03
Visual	1	39	0.91	0.008

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

Table 13

Summary of Shapiro-Wilk Tests for GJT Data on Items of Grammaticality Across Groups and in Each Individual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Grammaticality				
GR				
Pretest	4	0.90	77	<0.001
Auditory	2	0.86	38	<0.001
Visual	0	0.92	39	0.015
Posttest	3	0.74	77	<0.001
Auditory	2	0.75	38	<0.001
Visual	1	0.73	39	<0.001
UNG				
Pretest	0	0.96	77	0.15
Auditory	0	0.94	38	0.07
Visual	0	0.95	39	0.12
Posttest		0.96	77	0.16
Auditory	0	0.95	38	0.15
Visual	0	0.95	39	0.16

Note. GR = grammatical items; UNG = ungrammatical items.

Table 14

Summary of Shapiro-Wilk Tests for GJT Data on Items of Syntactic Structures
for Each Type of Grammaticality in Auditory Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	3	38	0.81	<0.001
Posttest	1	38	0.64	<0.001
TQ × UNG				
Pretest	0	38	0.89	0.001
Posttest	3	38	0.89	0.001
NA × GR				
Pretest	0	38	0.86	<0.001
Posttest	0	38	0.76	<0.001
NA × UNG				
Pretest	0	38	0.86	<0.001
Posttest	0	38	0.90	0.003
CON × GR				
Pretest	0	38	0.86	<0.001
Posttest	1	38	0.53	<0.001
CON × UNG				
Pretest	0	38	0.87	0.001
Posttest	0	38	0.89	0.001

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items; each syntactic structure: 8 items, and each type of grammaticality: 12 items (i.e., each type of syntactic structure includes 4 grammatical and 4 ungrammatical items, 24 in total).

Table 15

Summary of Shapiro-Wilk Tests for GJT Data on Items of Syntactic Structures
for Each Type of Grammaticality in Visual Group

		Shapiro-Wilk Test		
Variable	Outlier	df	*F*	*P*
TQ × GR				
Pretest	3	39	0.82	<0.001
Posttest	4	39	0.51	<0.001
TQ × UNG				
Pretest	0	39	0.90	0.002
Posttest	0	39	0.88	0.001
NA × GR				
Pretest	0	39	0.86	<0.001
Posttest	3	39	0.70	<0.001
NA × UNG				
Pretest	0	39	0.86	<0.001
Posttest	4	39	0.90	0.004
CON × GR				
Pretest	3	39	0.87	<0.001
Posttest	1	39	0.56	<0.001
CON × UNG				
Pretest	0	39	0.87	<0.001
Posttest	0	39	0.89	0.002

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items; each syntactic structure: 8 items, and each type of grammaticality: 12 items (i.e., each type of syntactic structure includes 4 grammatical and 4 ungrammatical items, 24 in total).

Goal 1. The goal is to examine how the mean improvements in performance between the pre- and posttests differed between the two interventions (auditory and visual), as well as whether any difference between the two, if present, was influenced by the type of syntactic structure or grammaticality.

As the descriptive statistics in Table 10 show, the overall pattern across all conditions indicated improvement following either of the modality-specific interventions, as reflected in the higher mean scores obtained in posttests compared to the pretests. Across the sample as a whole, the mixed ANOVA showed a significant main effect for the testing session (i.e., collapsing over all other variables), $F(1, 75) = 83.05$, $p < 0.001$, with a large effect size, partial eta-squared, $\eta p^2 = 0.52$ (i.e., based on Cohen's scale, 1988, 0.01 = small effect, 0.06 = moderate effect, 0.14 = large effect), reflecting an overall increase in the mean GJT score from the pretest ($M = 12.49$, $SE = 0.37$) to posttests ($M = 16.57$, $SE = 0.38$).

The overall pattern of the pre- and posttest data relating to different syntactic structures revealed a significant main effect of syntactic structure, $F(2, 150) = 17.66$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.19$, demonstrating that there was a difference in mean GJT scores across the three types of structure. Post hoc analysis with Bonferroni adjustment showed that the average scores were significantly higher in the TQ condition ($N = 8$, $M = 5.06$, $SE = 0.11$) than in either the NA ($N = 8$, $M = 4.10$, $SE = 0.13$) or the CON ($N = 8$, $M = 4.49$, $SE = 0.12$) conditions ($p < 0.001$ in both instances, $SE = 0.17$, $SE = 0.14$, respectively), whereas CON elicited the second highest average score, but it failed to be significantly higher than the NA scores ($p = 0.13$, $SE = 0.16$). Furthermore, the overall pattern of data pertaining to the distinct type of grammaticality indicated a significant main effect of grammaticality, $F(1, 75) = 160.07$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.68$, demonstrating that the overall performance in two pre- and posttests on grammatical items ($N = 12$, $M = 9.13$, $SE = 0.20$) is significantly better than ungrammatical items ($N = 12$, $M = 5.39$, $SE = 0.21$).

The interaction between the testing session and syntactic structure was not statistically significant, $F(2, 150) = 1.48$, $p = 0.22$, $\eta p^2 = 0.01$, demonstrating that the overall amount of learning did not vary considerably for different structures: TQ, NA, and CON ($M = 1.26$, $SE = 0.21$, $M = 1.26$, $SE = 0.22$, $M = 1.69$, $SE = 0.22$, respectively). These non-significant

interactions are visualised in the interaction plot shown in Figure 7 below, in which the first

slope (representing the TQ structure) and the last slope (representing the NA structure) have

very similar illustrations. Although the middle slope (representing the CON structure)

illustrates a fairly distinct effect compared to the other two slopes, this differential effect does

not reach a statistical significance.
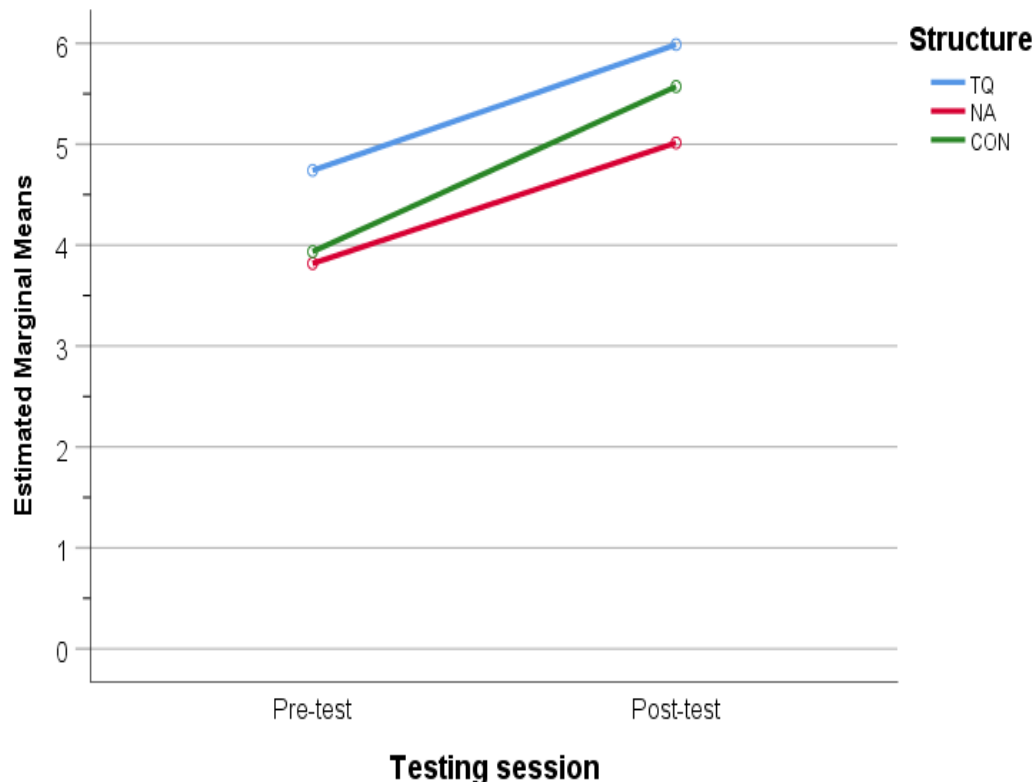
Figure 7. Performance for each structure type in the pre- and posttest across
both modality groups.

In a similar way, the analysis indicated that the interaction between the testing session

and grammaticality was not statistically significant, $F(1, 75) = 0.08$, $p = 0.76$, $\eta p^2 = 0.00$,

demonstrating that the overall improvement on grammatical items did not differ significantly

compared to ungrammatical items, as revealed by the parallel slopes of improvement on

grammatical and ungrammatical items illustrated in Figure 8 below.
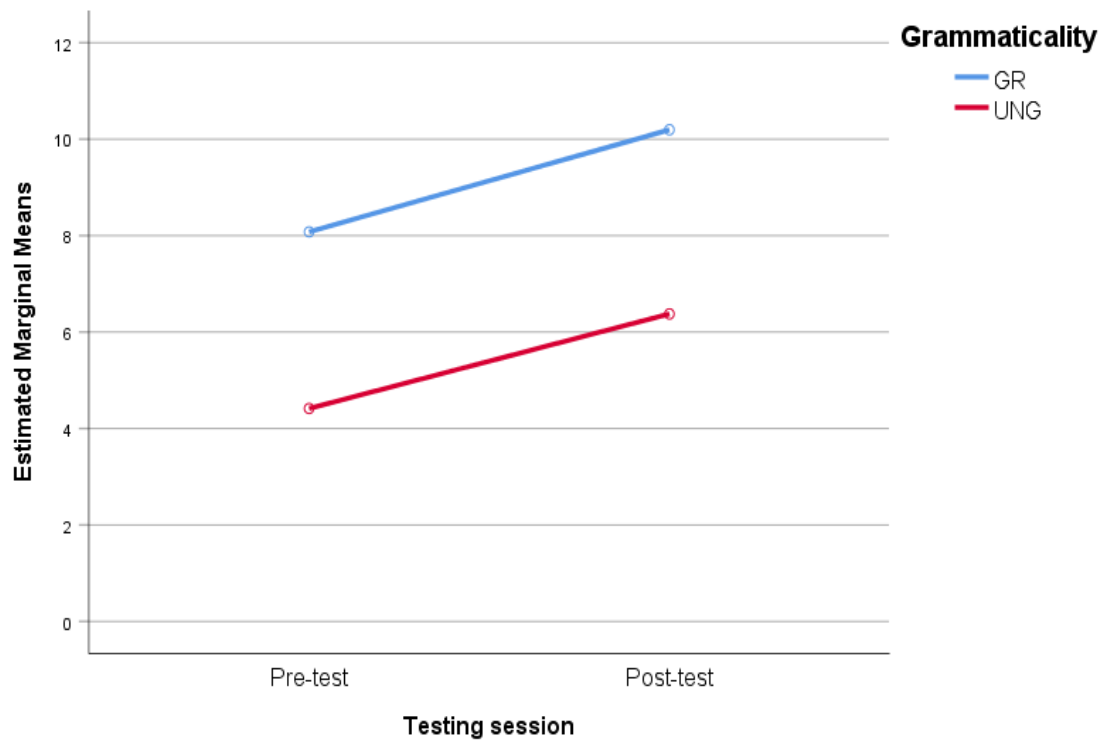
Figure 8. Performance for each grammaticality type in the pre- and posttest across both modality groups.

The descriptive statistics in Table 10 display the pattern of data relating to modality interventions indicating a greater mean improvement from the pre- to posttests in the visual intervention modality ($M = 4.94$, $SE = 0.64$) compared to the auditory intervention modality ($M = 3.18$, $SE = 0.61$). This pattern was present in all individual conditions (i.e., both items of syntactic structures and grammaticality). However, the ANOVA analysis revealed no significant interaction between testing session and modality, but was very close to being statistically significant, $F (1, 75) = 3.90$, $p = 0.052$, with a small effect size, $\eta p^2 = 0.05$, suggesting that performance did not significantly improve to a greater extent in the visual intervention compared to the auditory intervention, as supported by the partial overlap in Figure 9 below.
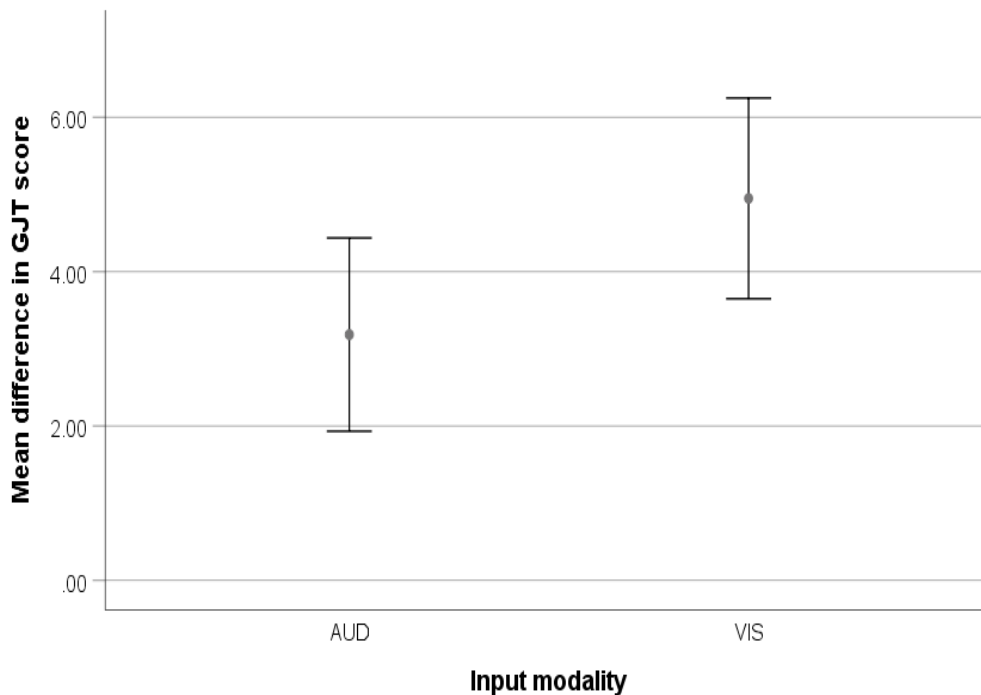
Figure 9. The difference in mean GJT improvement between auditory and visual modalities. Error bars indicate 95% confidence intervals.

However, because the difference in improvement from the pre- and posttests between the two modality interventions was very slightly below chance, determining whether the type of syntactic structure or grammaticality played a distinct role in this possible difference was worth mentioning. Deeper analysis did not indicate whether syntactic structure or grammaticality had a different effect on the possible difference between the two modality interventions, $F(2, 150) = 0.45$, $p = 0.63$, $\eta p^2 = 0.006$, $F(1, 75) = 1.99$, $p = 0.16$, $\eta p^2 = 0.02$, respectively.

Goal 2. The goal is to investigate whether auditory intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality.

The results first indicated that there was a significant main effect of testing session, $F(1, 37) = 26.51$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.41$, reflecting an increase in the mean GJT score in the auditory group from the pre- ($M = 12.76$, $SE = 0.45$) to posttests ($M =$

15.97, *SE* = 0.54). However, the analysis revealed neither a significant interaction between the testing session and syntactic structure, $F(2, 74) = 0.46$, $p = 0.63$, $\eta p^2 = 0.01$, nor the testing session and grammaticality, $F(1, 37) = 0.55$, $p = 0.46$, $\eta p^2 = 0.01$. Furthermore, and more importantly, the ANOVA revealed that there was no significant interaction between the testing session, grammaticality, and structure, $F(2, 74) = 2.23$, $p = 0.11$, $\eta p^2 = 0.05$, indicating that participants' improvement in performance on each of the different structures was not influenced by grammaticality.

Goal 3. The goal is to investigate whether visual intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality.

In a similar vein, the results revealed there was a significant main effect of the testing session, $F(1, 38) = 59.31$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.61$, indicating a positive change in the mean GJT score in the visual group from the pre- ($M = 12.07$, $SE = 0.58$) to posttests ($M = 17.28$, $SE = 0.54$). However, the analysis revealed neither a significant interaction between testing session and syntactic structure, $F(2, 76) = 1.58$, $p = 0.21$, $\eta p^2 = 0.04$, nor testing session and grammaticality, $F(1, 38) = 1.64$, $p = 0.20$, $\eta p^2 = 0.04$. Furthermore, and more importantly, the ANOVA revealed there was no significant interaction between the testing session, grammaticality, and structure, $F(2, 76) = 0.25$, $p = 0.77$, $\eta p^2 = 0.00$, indicating that participants' improvement in performance on each of the different structures was not influenced by grammaticality.

To summarise the results of analysis relating to the timed GJT, both auditory and visual exposure to L2 English syntax were found to result in successful acquisition of implicit grammatical knowledge. Although the results indicated no differentiable learning effects between visual and auditory interventions, the difference was very close to being statistically significant ($p = 0.052$). This suggests there was an apparent tendency for the process of

automating L2 grammatical knowledge to be fostered via the visual modality more successfully than the auditory modality. For performance on syntactic structure across the whole sample, no differing amounts of learning occurred for different syntactic structures, where similar results were present within each modality group. The results further showed that the mean improvement score in performance between the pre- and posttests was comparable in the case of both grammatical and ungrammatical items. Similar results were further present within each modality group. Finally, participants' improvement in performance on each of the different structures was not influenced by grammaticality within each modality group (summary tables of the results are shown altogether with EIT's results under the analysis of EIT results section).

4.4.2   Analysis of EIT results

Table 16 shows participants' mean raw scores, with standard errors in both the pre- and posttests in each of the two modality-specific interventions, including overall performance and means at each level of the syntactic construction and grammaticality conditions.

Table 16

Correct Mean (M) Raw EIT Scores with Standard Errors (SEs)

Condition	Pretest		Posttest		Improvement	
	M	*SE*	*M*	*SE*	*M*	*SE*
Overall performance	6.18	0.33	8.95	0.46	2.76	0.33
Auditory	5.95	0.44	8.13	0.66	2.18	0.50
Visual	6.41	0.50	9.74	0.62	3.33	0.42
Syntactic structure						
TQ	3.22	0.15	3.32	0.19	0.10	0.19
Auditory	3.21	0.21	3.21	0.29	0.00	0.31
Visual	3.23	0.22	3.44	0.26	0.20	0.23
NA	2.30	0.19	4.04	0.20	1.74	0.19
Auditory	2.29	0.25	3.76	0.30	1.47	0.30
VIS	2.31	0.28	4.31	0.25	2.00	0.22
CON	0.66	0.10	1.58	0.20	0.92	0.18
Auditory	0.45	0.13	1.16	0.29	0.71	0.25
Visual	0.87	0.15	2.00	0.26	1.12	0.25
Grammaticality						
GR	5.06	0.29	6.62	0.32	1.55	0.22
Auditory	4.74	0.38	5.95	0.45	1.21	0.31
Visual	5.38	0.43	7.28	0.43	1.89	0.31
UNG	1.12	0.12	2.32	0.22	1.20	0.20
Auditory	1.21	0.16	2.18	0.31	0.97	0.29
Visual	1.03	0.19	2.46	0.31	1.43	0.28

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items. The maximum raw score is 24 for overall performance, 8 for each syntactic structure, and 12 for each type of grammaticality (i.e., each type of syntactic structure includes four grammatical items and four ungrammatical items, for a total of 24).

Before performing inferential analyses to address the research questions, an independent-samples t-test was performed on the pretest scores to rule out differences between the auditory and visual groups prior to the intervention. The analysis demonstrated no significant difference in the EIT scores for auditory modality ($M = 5.95$, $SE = 0.44$) and visual modality ($M = 6.41$, $SE = 0.50$) groups, indicating that the sample was homogeneous

and that any differences observed in the posttests could be attributable to the modality intervention. In addition, it is also important to check the data of each variable used in the ANOVAs below in terms of the normality of distribution and outliers. The results of the Shapiro-Wilks tests revealed that the data of the variables that pertained to answering the first research question were normally distributed (i.e., the pre- and posttest performances for each group were checked individually). However, one outlier was observed in the posttest for the auditory group. The following tables: Table 17, Table 18, Table 19, Table 20, and Table 21 shows a summary of the data of all variables in terms of normality of distribution and outliers.

Table 17

Summary of Shapiro-Wilk Tests for EIT Data Across Both Modality Groups and in Each Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Overall performance				
Pretest	0	77	0.97	0.20
Posttest	0	77	0.97	0.17
Modality				
Auditory				
Pretest	0	38	0.95	0.14
Posttest	1	38	0.96	0.29
Visual				
Pretest	0	39	0.98	0.81
Posttest	0	39	0.97	0.39

Table 18

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures across Groups and Each Individual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Syntactic structure				
TQ				
Pretest	0	77	0.94	0.003
Auditory	0	38	0.93	0.027
Visual	2	39	0.92	0.015
Posttest	0	77	0.94	0.003
Auditory	0	38	0.94	0.050
Visual	3	39	0.94	0.058
NA				
Pretest	0	77	0.92	< 0.001
Auditory	0	38	0.93	0.030
Visual	0	39	0.92	0.011
Posttest	0	77	0.95	0.011
Auditory	0	38	0.94	0.065
Visual	0	39	0.93	0.028
CON				
Pretest	4	77	0.71	< 0.001
Auditory	2	38	0.60	< 0.001
Visual	0	39	0.78	< 0.001
Posttest	0	77	0.82	< 0.001
Auditory	1	38	0.70	< 0.001
Visual	0	39	0.90	0.003

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

Table 19

Summary of Shapiro-Wilk Tests for EIT Data on Items of Grammaticality Across Groups and in Each individual Group

| Variable | Outlier | Shapiro-Wilk Test | | |
		df	*F*	*P*
Grammaticality				
GR				
Pretest	0	77	0.96	0.041
Auditory	0	38	0.95	0.17
Visual	0	39	0.96	0.21
Posttest	0	77	0.96	0.051
Auditory	0	38	0.95	0.10
Visual	0	39	0.95	0.13
UNG				
Pretest	0	77	0.81	< 0.001
Auditory	0	38	0.84	< 0.001
Visual	3	39	0.76	< 0.001
Posttest	1	77	0.88	< 0.001
Auditory	1	38	0.82	< 0.001
Visual	0	39	0.90	0.004

Note. GR = grammatical items; UNG = ungrammatical items.

Table 20

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures for Each Type of Grammaticality in Auditory Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	2	38	0.90	0.002
Posttest	0	38	0.89	0.001
TQ × UNG				
Pretest	1	38	0.81	< 0.001
Posttest	3	38	0.81	< 0.001
NA × GR				
Pretest	0	38	0.90	0.003
Posttest	0	38	0.85	< 0.001
NA × UNG				
Pretest	0	38	0.65	< 0.001
Posttest	4	38	0.80	< 0.001
CON × GR				
Pretest	2	38	0.60	<0.001
Posttest	0	38	0.70	<0.001
CON × UNG				
Pretest	4	38	0.45	< 0.001
Posttest	4	38	0.45	< 0.001

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items; each syntactic structure: 8 items, and each type of grammaticality: 12 items (i.e., each type of syntactic structure includes 4 grammatical and 4 ungrammatical items, 24 in total).

Table 21

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures
for Each Type of Grammaticality in Visual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	0	39	0.88	0.001
Posttest	0	39	0.89	0.001
TQ × UNG				
Pretest	1	39	0.78	< 0.001
Posttest	3	39	0.78	< 0.001
NA × GR				
Pretest	0	39	0.88	0.001
Posttest	0	39	0.79	< 0.001
NA × UNG				
Pretest	5	39	0.50	< 0.001
Posttest	1	39	0.87	< 0.001
CON × GR				
Pretest	0	39	0.77	< 0.001
Posttest	0	39	0.88	0.001
CON × UNG				
Pretest	0	39	0.63	< 0.001
Posttest	0	39	0.63	< 0.001

Note. TQ = tag questions; NA = negative adverb; CON =
conditional; GR = grammatical items; UNG = ungrammatical
items; each syntactic structure: 8 items, and each type of
grammaticality: 12 items (i.e., each type of syntactic structure
includes 4 grammatical and 4 ungrammatical items, 24 in total).

Goal 1. The goal is to examine how the mean improvements in performance between

the pre- and posttests differed between the two interventions (auditory and visual), as well as

whether any difference between the two, if present, was influenced by the type of syntactic

construction or grammaticality.

As the descriptive statistics in Table 16 illustrate, the overall pattern across all

conditions indicated improvement, as reflected in the higher mean scores obtained in

posttests compared to the pretests. Across the sample as a whole, the mixed ANOVA showed a significant main effect for the testing session, $F(1, 75) = 70.83$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.48$, reflecting an overall increase in the mean EIT score from the pretest ($M = 6.18$, $SE = 0.33$) to the posttest ($M = 8.95$, $SE = 0.45$).

The overall pattern of the pre- and posttest data relating to different syntactic structures revealed a significant main effect of syntactic structure, $F(2, 150) = 123.29$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.62$, demonstrating that there was a difference in mean EIT scores across the three types of structure. Post hoc analysis with Bonferroni adjustment showed that the average scores in the TQ and NA conditions reflected similar patterns of performance ($p = 1.00$, $SE = 0.17$) while CON scores were significantly lower than the two other conditions ($p < 0.001$ in both instances, $SE = 0.13$, $SE = 0.14$, respectively). Furthermore, the overall pattern of data pertaining to the distinct type of grammaticality indicated a significant main effect of grammaticality, $F(1, 75) = 248.17$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.76$, demonstrating that the overall performance in the two pre- and posttests on grammatical items ($M = 5.83$, $SE = 0.28$) is significantly better than ungrammatical items ($M = 1.72$, $SE = 0.14$).

More importantly, the interaction between the testing session and syntactic structure was statistically significant, $F(2, 150) = 18.57$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.19$, indication that the overall amount of learning varied for different structures: TQ, NA, and CON. Deeper analysis from the same mixed-ANOVA revealed that the first contrast comparing improvement in performance on the NA structure ($M = 1.73$, $SE = 0.19$) to TQ structure ($M = 0.10$, $SE = 0.19$, the TQ was used as a baseline category), indicated a significant difference $F(1, 76) = 32.25$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.29$, demonstrating that participants' level of improvement based on learning was greater in the NA structure than in the TQ structure. The second contrast comparing improvement in

performance on the CON structure to TQ structure demonstrated a significant difference $F(1, 76) = 11.02$, $p = 0.001$, with a large effect size, $\eta p^2 = 0.12$, indicating that participants' level of improvement based on learning was greater in the CON structure ($M = 0.92$, $SE = 0.18$) than in the TQ structure. The differential improvements are visualised in the interaction plot shown in Figure 10 below, in which the slope (representing the TQ structure) illustrates a distinct effect compared to NA and CON slopes. This illustration makes it clear that the pattern of improvement between the pre- and posttests was significantly stronger in the NA and CON condition than TQ condition; the latter case elicited a very week pattern of improvement.
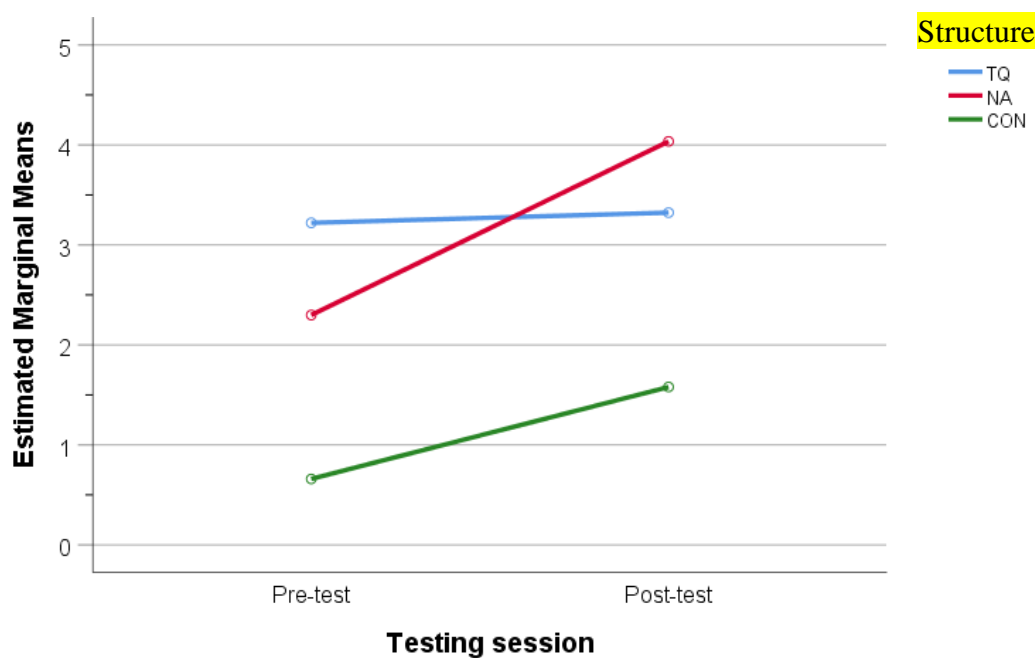
Figure 10. Performance for each structure type in the pre- and posttest across both modality groups.

In a similar way, the analysis indicated that the interaction between the testing session and grammaticality was not statistically significant, $F(1, 75) = 1.61$, $p = 0.20$, $\eta p^2 = 0.02$, demonstrating that the overall improvement in performance on grammatical items did not differ significantly compared to ungrammatical items, as supported by the parallel slopes of improvement in grammatical and ungrammatical items illustrated in Figure 11 below.
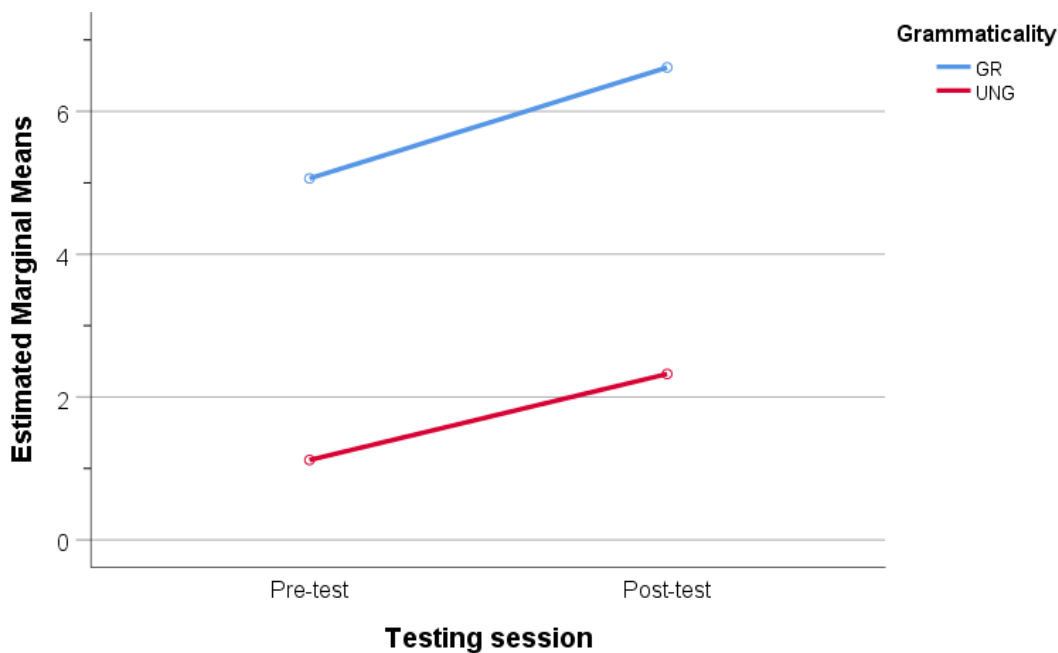
Figure 11. Performance for each grammaticality type in the pre- and posttest across both modality groups.

The descriptive statistics in Table 16 further display the pattern of data relating to modality interventions indicating a greater mean improvement from the pre- to posttests in the visual intervention modality ($M = 3.33$, $SE = 0.42$) compared to the auditory intervention modality ($M = 2.18$, $SE = 0.50$). This pattern was present in all individual conditions (i.e., both items of syntactic structures and grammaticality). The ANOVA analysis did not reveal a significant interaction between testing session and modality but was close to being statistically significant, $F(1, 75) = 3.07$, $p = 0.08$, $\eta p^2 = 0.03$, suggesting that difference in

performance on improvement did not significantly differ between the two modality

interventions, as supported by the partial overlap in Figure 12 below.
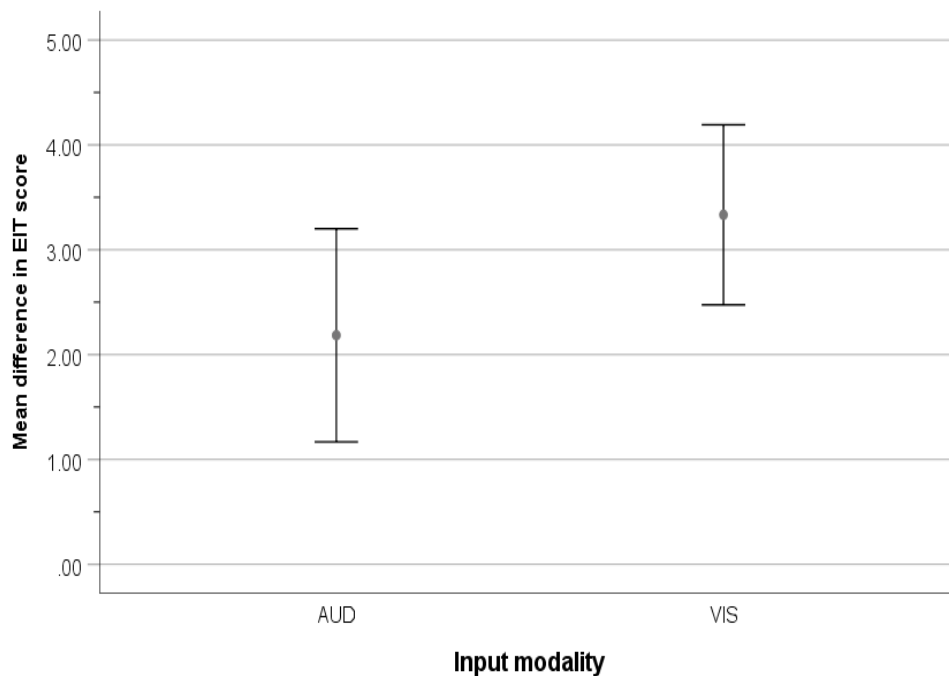
Figure 12. The difference in mean GJT improvement between auditory and visual
modalities. Error bars indicate 95% confidence intervals.

However, because the difference in improvement from the pre- and posttests between

the two modality interventions was slightly above chance, determining whether the type of

syntactic construction or grammaticality played a distinct role in this possible difference was

worth mentioning. Deeper analysis did not indicate whether syntactic construction or

grammaticality had a different effect on the possible difference between the two modality

interventions, $F(2, 150) = 0.18$, $p = 0.83$, $\eta p^2 = 0.002$, $F(1, 75) = 0.16$, $p = 0.68$, $\eta p^2 = 0.002$,

respectively.

Goal 2. The goal is to investigate whether auditory intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality.

The results indicated that there was a significant main effect of testing session, $F(1, 37) = 18.02$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.32$, indicating a positive change in the mean EIT score in the auditory group from the pre- ($M = 5.95$, $SE = 0.44$) to posttests ($M = 8.13$, $SE = 0.66$). Moreover, the ANOVA analysis revealed that there was a significant interaction between the testing session and syntactic structure, $F(2, 74) = 7.15$, $p = 0.001$, with a large effect size, $\eta p^2 = 0.16$, indicating that the overall amount of learning varied for different structures over data from the auditory group. Deeper analysis from the same ANOVA revealed that the first contrast comparing improvement in performance on the NA structure ($M = 1.47$, $SE = 0.30$) to TQ structure ($M = 0.00$, $SE = 0.31$, the TQ was used as a baseline category), indicated a significant difference $F(1, 37) = 9.55$, $p = 0.004$, with a large effect size, $\eta p^2 = 0.20$, demonstrating that participants' level of improvement based on learning was greater in the NA structure than in the TQ structure. The second contrast comparing improvement in performance on the CON structure ($M = 0.71$, $SE = 0.25$) to TQ structure demonstrated no significant difference, $F(1, 37) = 3.28$, $p = 0.07$, $\eta p^2 = 0.08$. The differential improvements are visualised in the interaction plot shown in Figure 13 below.
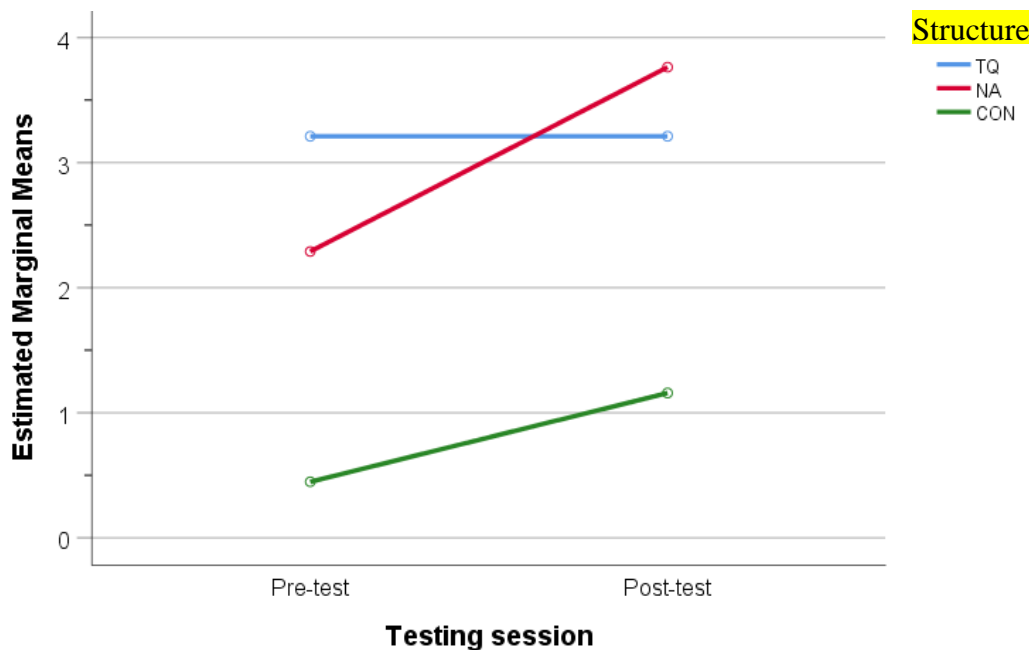
Figure 13. Performance for each structure type in the pre- and posttest by the auditory group.

However, the analysis revealed no significant difference between testing session and grammaticality, $F(1, 37) = 0.45$, $p = 0.50$, $\eta p^2 = 0.01$, indicating that the improvement in performance between the pre- and posttests was comparable in the case of both grammatical and ungrammatical items. More importantly, the analysis revealed that there was no significant interaction between testing session, grammaticality and structure, $F(2, 74) = 2.41$, $p = 0.09$, $\eta p^2 = 0.06$, indicating that participants' improvement in performance on each of the different structures was not influenced by grammaticality.

Goal 3. The goal is to investigate whether visual intervention participants' improvement in performance on each of the different syntactic structures was influenced by grammaticality.

In a similar vein, the results indicated there was a significant main effect of the testing session, $F(1, 38) = 60.62$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.61$, reflecting an increase in the mean EIT score for the visual group from the pre- ($M = 6.41$, $SE = 0.50$) to posttests ($M = 9.74$, $SE = 0.62$). Furthermore, the ANOVA analysis revealed that there was a significant

interaction between the testing session and syntactic structure, $F(2, 78) = 16.08$, $p = 0.001$, with a large effect size, $\eta p^2 = 0.29$, indicating that the overall amount of learning varied for different structures in the visual group. Deeper analysis from the same ANOVA revealed that the first contrast comparing improvement in performance on the NA structure ($M = 2.00$, $SE = 0.22$) to TQ structure ($M = 0.20$, $SE = 0.23$, the TQ was used as a baseline category) demonstrated that the pattern of improvements on the two structures differed significantly $F(1, 38) = 29.04$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.43$, suggesting that the improvement in performance on NA structure is greater than TQ structure. The second contrast comparing improvement in performance on the CON structure ($M = 1.12$, $SE = 0.25$) to TQ structure, indicated a significant difference, $F(1, 38) = 9.01$, $p = 0.005$, with a large effect size, $\eta p^2 = 0.19$, suggesting that participants' level of improvement based on learning was greater in the CON structure than in the TQ structure. These interactions are visualised in the interaction plot shown in Figure 14 below.
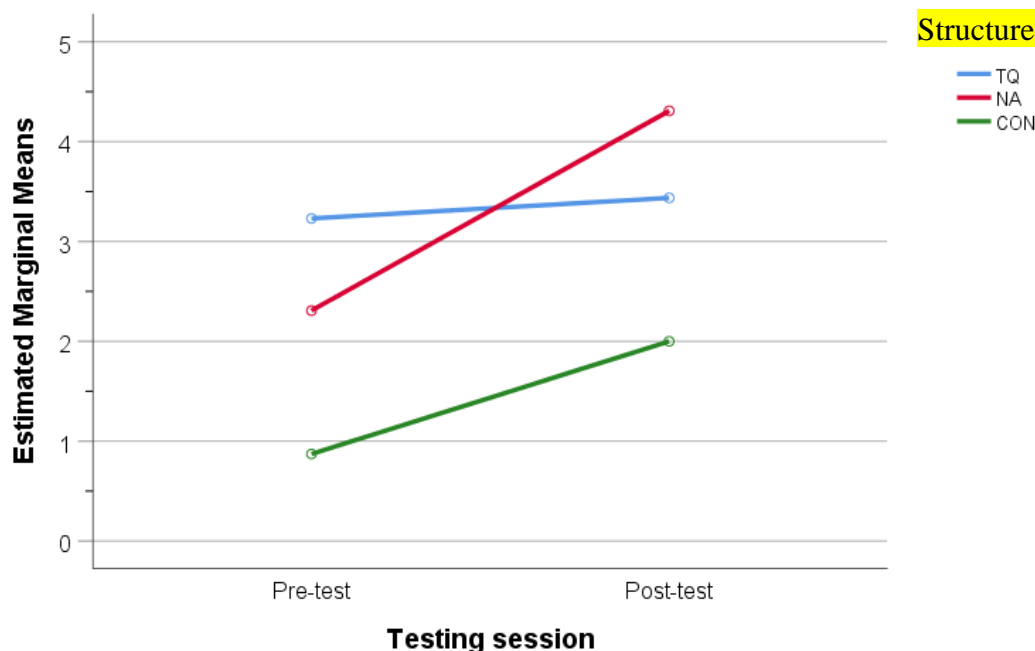
Figure 14. Performance for each structure type in the pre- and posttest by the visual group.

In relation to the interaction between testing session and grammaticality, the analysis revealed no significant difference between the two variables, , $F(1, 38) = 1.19$, $p = 0.28$, $\eta p^2 = 0.03$, suggesting that the improvement in performance between the pre- and posttests was not different between grammatical and ungrammatical items.

However, and more importantly, the analysis revealed that there was a significant interaction between the testing session, grammaticality and structure, $F(2, 76) = 4.06$, $p = 0.02$, with a moderate effect size, $\eta p^2 = 0.09$, indicating that participants' improvement in performance on each of the different structures was influenced by grammaticality. Deeper analysis from the same ANOVA revealed that the first contrast comparing improvement in performance on the NA structure to TQ structure (the TQ was used as a baseline category) did not differ significantly between grammatical and ungrammatical items, $F(1, 38) = 0.36$, $p = 0.55$, $\eta p^2 = 0.00$. The second contrast comparing improvement in performance on the CON structure to TQ structure differed significantly between grammatical and ungrammatical items, $F(1, 38) = 7.97$, $p = 0.007$, with a large effect size, $\eta p^2 = 0.17$. These interactions are visualised in the interaction plot shown in Figure 15 and 16 below, in which the slopes of the lines representing the TQ and NA structures in grammatical results were comparable to a similar extent for the same structures in ungrammatical results (i.e., both are crossed out). This is unlike the slopes representing the CON and TQ in the grammatical results which were different from those in the ungrammatical items.
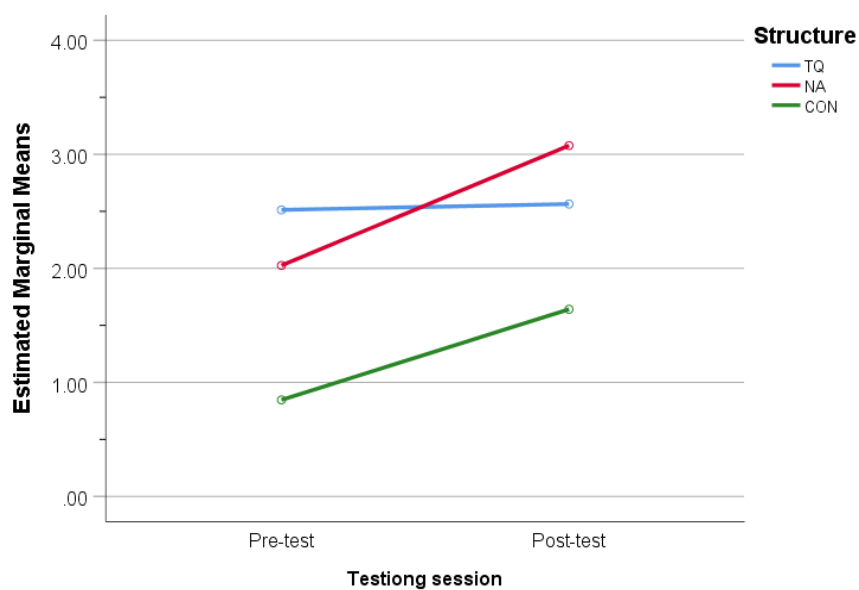
Figure 15. Performance for each structure type on grammatical items in the pre- and posttest by the visual group.
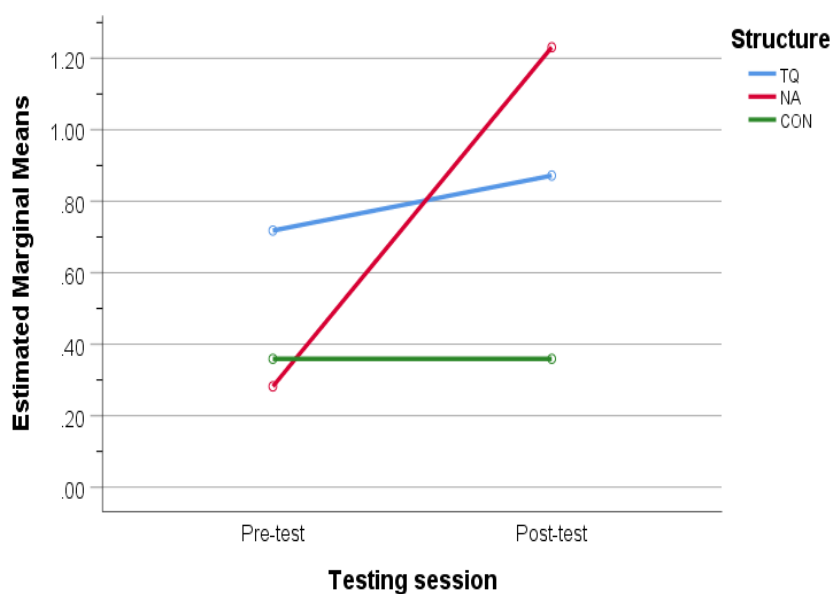
Figure 16. Performance for each structure type on ungrammatical items in the pre- and posttest by the visual group.

To summarise the results of analysis pertaining to the EIT, both auditory and visual exposure to L2 English syntax were found to result in successful acquisition of implicit grammatical knowledge. The results indicated no differentiable learning effects between visual and auditory interventions. However, the difference was fairly close to being statistically significant ($p = 0.08$). This suggests there was an apparent tendency for the process of automating L2 grammatical knowledge to be fostered via the visual modality more successfully than the auditory modality. For performance on syntactic structure across the whole sample, amounts of learning varied for different syntactic structures, where the mean improvement in TQ was significantly lower than that observed in NA and CON structures. Similar results were present in the visual group. In the auditory group, however, the difference lay between TQ and NA structures. The results further showed that the improvement in performance between the pre- and posttests was comparable in the case of both GR and UNG items. Similar results were further present within each modality group. Finally, participants' improvement in performance on each of the different structures was influenced by grammaticality for only visual group.

Table 22

Summary of Mixed-ANOVAs for GJT and EIT Data Across Both Modality Groups

Variables	GJT			EIT		
	df	*F*	*P*	df	*F*	*P*
Main effects						
Testing session	1, 75	83.5	< 0.001	1, 75	70.83	< 0.001
Syntactic structure	2, 150	17.66	< 0.001	2, 150	123.29	< 0.001
Grammaticality	1, 75	160.07	< 0.001	1, 75	248.17	< 0.001
Interaction effects						
Testing × structure	2, 150	1.48	0.22	2, 150	18.57	< 0.001
Testing × grammaticality	1, 75	0.08	0.76	1, 75	1.61	0.20
Testing × modality	1, 75	3.90	0.052	1, 75	3. 07	0.08

Table 23

Summary of Repeated Measures ANOVAs for GJT and EIT Data in Auditory Group

Variables	GJT			EIT		
	df	*F*	*P*	df	*F*	*P*
Main effects						
Testing session	1, 37	26.51	< 0.001	1, 37	18.02	< 0.001
Interaction effects						
Testing × structure	2, 74	0.46	0.63	2, 74	7.15	< 0.001
Testing × grammaticality	1, 37	0.55	0.46	1, 37	1.97	0.16
Testing × structure × grammaticality	2, 74	2.23	0.11	2, 74	2.41	0.09

Table 24

Summary of Repeated Measures ANOVAs for GJT and EIT Data in Visual Group

Variables	GJT			EIT		
	df	*F*	*P*	df	*F*	*P*
Main effects						
Testing session	1, 38	59.31	< 0.001	1, 38	60.62	< 0.001
Interaction effects						
Testing × structure	2, 76	1.58	0.21	2, 78	16.08	< 0.001
Testing × grammaticality	1, 38	1.64	0.20	1, 38	3.80	= 0.28
Testing × structure × Structure	2, 76	0.25	0.77	2, 76	4.40	< 0.01

4.4.3 Working memory analysis

The average score of performance across the two experimental conditions on the OSPAN (a measure of WM capacity) was 66.03 (*SD* = 6. 41, *SE* = 0.73) out of a possible total of 75. Participants in the auditory group elicited an average of 66.50 (*SD* = 6.44, *SE* = 1.04), while those in the visual group scored an average of 65.56 (*SD* = 6.43, *SE* = 1.03). In the digit span (a measure of phonological short-term memory), the mean score across the two experimental conditions was 6.34 (*SD* = 1.42, *SE* = 0.16) out of a possible total of 9. The mean score among the auditory group was 6.13 (*SD* = 1.49, *SE* = 0.24), while the mean score

among the visual group was 6.54 ($SD = 1.33$, $SE = 0.21$). A Mann-Whitney U Test revealed that there was no significant difference in OSPAN scores between the auditory group ($Md = 67$) and the visual group ($Md = 68$), $U = 666.500$, $z = -0.76$, $p = 0.44$. Similarly, there was no significant difference in scores on the digit span measure between participants in the auditory group ($Md = 6$) and those in the visual group ($Md = 6$), $U = 900.500$, $z = 1.67$, $p = 0.09$. Spearman's rank-order correlation coefficient was calculated over all participants and revealed no significant correlation between OSPAN and digit span performance, $r = 0.09$, $p = 0.41$. There was also no correlation when this test was carried out specifically among participants in the auditory group, $r = 0.04$, $p = 0.77$. However, there was a weak-to-moderate positive correlation between OSPAN and DS scores among participants in the visual group, $r = .21$, $p = 0.18$.

Regression analysis by GJT across the groups

A simple linear regression was performed to investigate whether WM capacity (as measured by OSPAN task) was a significant predictor of overall improvement in performance on the timed GJT across the two experimental modality groups (auditory and visual). The results indicated that WM was not a significant predictor of improvement in the timed GJT performance ($p = 0.83$). The model explained 0.1 % of the variance and was not able to predict the improvement in GJT scores to a significant extent, $F(1, 76) = 0.04$, $p = 0.83$.

To explore whether WM predicts posttest performance in the GJT when pretest and IELTS (International English Language Testing System) scores are controlled, a multiple linear regression was conducted. The results indicated that none of variables: pretest, WM, IELTS score: 6.0, and IELTS score: 6.5, was a significant predictor of posttest performance. The model explained 18.4% of the variance and was able to predict posttest scores to a significant extent, $F(4, 76) = 4.05$, $p = 0.005$. Predicted posttest performance was equal to -

0.71 + 0.20 (pretest score) + 0.15 (WM score) + 0.47 (IELTS score: 6.0) + 0.95 (IELTS score: 6.5), where pretest and WM scores were measured in points, and IELTS scores were coded into three dummy variables: 6.5, 6.0, and 5.5 where the last band score was treated as a reference point. As participants' pretest performance on the GJT increased by one point, their posttest performance increased by 0.20 points. As WM capacity increased by one point, the model's predicted GJT score in the posttest increased by 0.15 points. A participant with an IELTS score of 6.0 was predicted to earn 0.47 more points on the GJT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 0.95 more points on the GJT than a participant with an IELTS score of 5.5.

Table 25

Summary of Predictors of PostTest Performance in GJT Across the Two Groups

	Model	
Variable	**Est.**	**SE**
Constant	– 0.71	(0.53)
Pretest	0.20†	(0.11)
WM	0.15	(0.11)
IELTS (ref. 5.5)		
Score 6.0	0.47	(0.56)
Score 6.5	0.95†	(0.55)
R^2	0.13	

Note. n = 77; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; † p < 0.10.

Regression analysis by GJT for auditory group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of improvement in performance on all GJT items for auditory group. The results indicated that WM was a significant predictor of improvement in the timed GJT

performance ($p = 0.015$). The model explained 13 % of the variance and was able to predict the improvement in GJT scores to a significant extent, $F(1, 37) = 6.50$, $p = 0.015$. Predicted improvement in GJT performance was equal to $-0.25 + 0.37$ (WM score), improvement in GJT scores and WM scores were measured in points. As WM capacity increased by one point, the model's predicted improvement in GJT increased by 0.37 points.

Due to the significant correlation observed between WM capacity and accuracy on GJT among participants in the auditory group, further analysis by performing several simple regressions was necessary to explore which type of grammaticality and syntactic structure drives this correlation. The results indicated that WM was not a significant predictor of improvement in performance on grammatical items ($p = 0.47$), TQ structure ($p = 0.84$), and CON structure ($p = 0.25$), where all models were not able to predict the improvement in GJT to a significant extent, $F(1, 37) = 0.51$, $p = 0.47$, $F(1, 37) = 0.04$, $p = 0.84$ and $F(1, 37) = 1.34$, $p = 0.25$, respectively. For grammatical items, the model explained - 1 % of the variance for grammatical items, - 2 % of the variance for items of TQ structure, and 0.9 % of the variance for items of CON structure.

However, the results indicated that WM was a significant predictor of improvement in performance on ungrammatical items and items of NA structure, $F(1, 37) = 7.64$, $p = 0.009$ and $F(1, 37) = 12.75$, $p = 0.001$, respectively. For the relationship between WM capacity and improvement in performance on ungrammatical items, the model explained 15 % of the variance, where the predicted improvement was equal to $-0.63 + 0.40$ (WM score). As WM capacity increased by one point, the model's predicted improvement in GJT increased by 0.40 points. The model for the relationship between WM capacity and improvement in performance on items of NA structure explained 24 % of the variance, where the predicted improvement was equal to $-0.11 + 0.55$ (WM score). As WM capacity increased by one

point, the model's predicted improvement in NA structure increased by 0.55 points. See

Table 26 below for the summary of all relationships.

Table 26

Summary of Relationships Between WM and Change from the Pre- to Posttests in GJT for Auditory Group in Overall Performance and in Individual Conditions

Variable	Overall		UNG		NA		GR		TQ		CON	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Constant	- 0.25	0.14	- 0.63	0.14	- 0.11	0.15	- 0.26	0.17	- 0.19	0.15	- 0.20	0.15
WM	0.37*	0.14	0.40**	0.14	0.55**	0.15	0.12	0.17	0.03	0.15	0.18	0.15
R^2	0.13		0.15		0.24		- 0.01		- 0.02		0.009	

Note. $n = 38$; WM = working memory; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; ** $p \leq$ 0.01; * $p \leq 0.05$; TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items.

To find out to which extent the strength of observed significant correlations above (i.e., WM-overall, WM-UNG, and WM-NA) were affected when IELTS scores (i.e., 6.0 and 6.5) were controlled, three multiple linear regressions were conducted.

The first multiple regression was carried out to investigate whether pretest, International English Language Testing System (IELTS) scores (IELTS score: 6.0 and 6.5), and WM were significant predictors of posttest performance in the GJT. The results indicated that only WM was a significant predictor of posttest performance on all timed GJT items ($p = 0.006$), while IELTS score 6.0 and 6.5, and pretest were not. The model explained 35 % of the variance and was able to predict posttest scores to a significant extent, $F(4, 37) = 6.06, p = 0.001$. Predicted posttest performance was equal to $- 0.67 + 0.12$ (pretest score) + 0.39 (WM score) + 0.01 (IELTS score: 6.0) + 0.78 (IELTS score: 6.5), where pretest and WM scores were measured in points, IELTS scores were coded into three dummy variables: 6.5, 6.0, and 5.5 where the last band score (5.5) was treated as a reference point. As participants'

pretest performance on the GJT increased by one point, their posttest performance increased by 0.12 points. As WM capacity increased by one point, the model's predicted GJT score in the posttest increased by 0.39 points. A participant with an IELTS score of 6.0 was predicted to earn 0.01 points on the GJT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 0.78 more points on the GJT than a participant with an IELTS score of 5.5.

Table 27

Summary of Predictors of Overall Posttest Performance on the GJT for Auditory Group

	Model 1	
Variable	**Est.**	**SE**
Constant	– 0.67	(0.56)
Pretest	0.12	(0.15)
WM	0.39**	(0.13)
IELTS (ref. 5.5)		
Score 6.0	0.01	(0.60)
Score 6.5	0.78	(0.59)
R^2	0.35	

Note. n = 38; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; ** p ≤ 0.01.

A second multiple linear regression was performed to examined whether pretest, IELTS scores, and WM were significant predictors of posttest performance on NA items. The results indicated that an IELTS score (6.5) and WM were significant predictors of posttest performance ($p = 0.039$, $p = 0.003$, respectively), while an IELTS score (6.0) and pretest were not. The model explained 45 % of the variance and was able to predict posttest scores to a significant extent, $F(4, 37) = 8.63$, $p < 0.001$. Predicted posttest performance was equal to – 0.86 + 0.15 (pretest score) + 0.41 (WM score) + 0.36 (IELTS score: 6.0) + 1.21 (IELTS

score: 6.5). As participants' pretest performance on the GJT increased by one point, their posttest performance increased by 0.15 points. As WM capacity increased by one point, the model's predicted GJT score in the posttest increased by 0.41 points. A participant with an IELTS score of 6.0 was predicted to earn 0.36 more points on the GJT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 1.21 more points on the GJT than a participant with an IELTS score of 5.5.

Table 28

Summary of Predictors of Posttest Performance on NA Items of EIT for Auditory Group

	Model 1	
Variable	Est.	SE
Constant	− 0.86	(0.52)
Pretest	0.15	(0.13)
WM	0.41**	(0.12)
IELTS (ref. 5.5)		
Score 6.0	0.36	(0.56)
Score 6.5	1.21	(0.56)
R^2	0.45	

Note. n = 38; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; ** p $\leq$ 0.01.

A third multiple linear regression was performed to examined whether pretest, IELTS scores, and WM were significant predictors of posttest performance on ungrammatical items. The results indicated that pretest and WM were significant predictors of posttest performance (p = 0.02, p = 0.02, respectively), while both IELTS scores: 6.0 and 6.5 were not. The model explained 42 % of the variance and was able to predict posttest scores to a significant extent, $F(4, 37)$ = 7.77, $p < 0.001$. Predicted posttest performance was equal to − 0.31 + 0.30 (pretest score) + 0.30 (WM score) − 0.26 (IELTS score: 6.0) + 0.66 (IELTS score: 6.5). As

participants' pretest performance on the GJT increased by one point, their posttest

performance increased by 0.30 points. As WM capacity increased by one point, the model's

predicted GJT score in the posttest increased by 0.30 points. A participant with an IELTS

score of 6.0 was predicted to lose 0.26 points on the GJT than a participant with an IELTS

score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 0.66 more points

Table 29

Summary of Predictors of Posttest Performance on Ungrammatical Items of EIT for Auditory group

	Model 1	
Variable	Est.	SE
Constant	– 0.31	(0.54)
Pretest	0.30	(0.12)
WM	0.30*	(0.12)
IELTS (ref. 5.5)		
Score 6.0	– 0.26	(0.57)
Score 6.5	0.66	(0.26)
R^2	0.42	

Note. n = 38; WM = working memory; IELTS =
International English Language Testing System;
Est. = estimate. SE = standard error. R^2 = adjusted
R-squared; * p $\leq$ 0.05.

on the GJT than a participant with an IELTS score of 5.5.

Regression analysis by GJT for visual group

A simple linear regression was performed to investigate whether WM capacity was a

significant predictor of overall improvement in performance on the timed GJT for visual

group. The results indicated that WM was not a significant predictor of improvement in the

timed GJT performance (p = 0.08). The model explained 5.4 % of the variance and was not

able to predict the improvement in GJT scores to a significant extent, $F(1, 38) = 3.17$, $p = 0.08$.

To explore whether WM predicts posttest performance in the GJT when pretest and IELTS scores are controlled, a multiple regression was conducted. The results of a multiple linear regression indicated that only pretest was a significant predictor of posttest performance ($p = 0.02$), while WM, IELTS score 0.6 and 6.5 were not. The model explained 16.8% of the variance and was unable to predict posttest scores to a significant extent, $F(4, 38) = 1.71$, $p = 0.16$.

As there was no significant relationship between WM and improvement in the GJT for the visual group, there was no need for deeper analysis for the relationships between WM and grammaticality items and items of syntactic structures.

Regression analysis by EIT across groups

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of improvement in performance on all items in the EIT across the two experimental groups (auditory and visual). The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.08$). The model explained 2.6 % of the variance and was not able to predict the improvement in EIT scores to a significant extent, $F(1, 76) = 2.99$, $p = 0.08$.

To explore whether WM predicts posttest performance in the EIT when pretest, digit span (DS, as a measure of phonological short-term memory), and IELTS scores: 6.0 and 6.5 are controlled, a multiple linear regression was conducted. The results indicated that only pretest was a significant predictor of posttest performance ($p < 0.001$), while WM, DS, IELTS scores were not. The model explained 49.6 % of the variance and was able to predict posttest scores to a significant extent, $F(5, 76) = 15.94$, $p < 0.001$. Predicted posttest performance was equal to - 0.43 + 0.69 (pretest score) - 0.15 (WM score) + 0.36 (DS score) +

0.29 (IELTS score: 6.0) + 0.57 (IELTS score: 6.5), where pretest, WM, and DS scores were

measured in points, and IELTS scores were coded into three dummy variables: 6.5, 6.0, and

5.5 where the last band score was treated as a reference point. As participants' pretest

performance on the EIT increased by one point, their posttest performance increased by 0.69

points. As WM capacity increased by one point, the model's predicted EIT score in the

posttest decreased by 0.15 points. As DS increased by one point, the model's predicted EIT

score in the posttest decreased by 0.36 points. A participant with an IELTS score of 6.0 was

predicted to earn 0.29 more points on the EIT than a participant with an IELTS score of 5.5.

A participant with an IELTS score of 6.5 was predicted to earn 0.57 more points on the EIT

than a participant with an IELTS score of 5.5.

Table 30

Summary of Predictors of Posttest Performance in EIT Across the Two Groups

Variable	Model 1	
	Est.	SE
Constant	- 0.43	(0.41)
Pretest	0.69***	(0.08)
WM	- 0.15	(0.08)
DS	0.36	(0.08)
IELTS (ref. 5.5)		
Score 6.0	0.29	(0.43)
Score 6.5	0.57	(0.43)
R^2	0.49	

Note. $n = 38$; WM = working memory; IELTS =
International English Language Testing System; Est.
= estimate. SE = standard error. R^2 = adjusted R-
squared; *** $p < 0.001$.

Regression analysis by EIT for auditory group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of overall improvement in performance on the EIT for auditory group. The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.15$). The model explained 2.9 % of the variance and was unable to predict the improvement in EIT scores to a significant extent, $F(1, 37) = 2.11$, $p = 0.15$.

To explore whether WM predicts posttest performance in the EIT when pretest, DS, and IELTS scores: 6.0 and 6.5 are controlled, a multiple linear regression was conducted. The results indicated that only pretest was a significant predictor of posttest performance ($p < 0.001$), while WM, DS, IELTS scores were not. The model explained 49.3 % of the variance and was able to predict posttest scores to a significant extent, $F(5, 37) = 8.18$, $p < 0.001$. Predicted posttest performance was equal to - 0.59 + 0.65 (pretest score) - 0.22 (WM score) + 0.05 (DS score) + 0.06 (IELTS score: 6.0) + 0.74 (IELTS score: 6.5). As participants' pretest performance on the EIT increased by one point, their posttest performance increased by 0.65 points. As WM capacity increased by one point, the model's predicted EIT score in the posttest decreased by 0.22 points. As DS increased by one point, the model's predicted EIT score in the posttest increased by 0.05 points. A participant with an IELTS score of 6.0 was predicted to earn 0.06 more points on the EIT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 0.74 more points on the EIT than a participant with an IELTS score of 5.5.

Table 31

Summary of Predictors of Posttest Performance in EIT for Auditory Group

Variable	Model 1	
	Est.	SE
Constant	- 0.59	(0.52)
Pretest	0.65***	(0.15)
WM	- 0.22†	(0.12)
DS	0.05	(0.11)
IELTS (ref. 5.5)		
Score 6.0	0.06	(0.54)
Score 6.5	0.74	(0.55)
R^2	0.49	

Note. n = 38; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; *** $p < 0.001$; † $p < 0.10$.

Regression analysis by EIT for visual group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of overall improvement in performance on the EIT for visual group. The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.15$). The model explained - 1 % of the variance and was unable to predict the improvement in EIT scores to a significant extent, $F(1, 38) = 0.63$, $p = 0.43$.

To explore whether WM predicts posttest performance in the EIT when pretest, DS, and IELTS scores: 6.0 and 6.5 are controlled, a multiple linear regression was conducted. The results indicated that only pretest was a significant predictor of posttest performance ($p < 0.001$), while WM, DS, IELTS scores were not. The model explained 49.3 % of the variance and was able to predict posttest scores to a significant extent, $F(5, 38) = 8.38$, $p < 0.001$. Predicted posttest performance was equal to - 0.27 + 0.64 (pretest score) - 0.09 (WM score) +

0.08 (DS score) + 0.39 (IELTS score: 6.0) + 0.43 (IELTS score: 6.5). As participants' pretest performance on the EIT increased by one point, their posttest performance increased by 0.64 points. As WM capacity increased by one point, the model's predicted EIT score in the posttest decreased by 0.09 points. As DS increased by one point, the model's predicted EIT score in the posttest increased by 0.08 points. A participant with an IELTS score of 6.0 was predicted to earn 0.39 more points on the EIT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to earn 0.43 more points on the EIT than a participant with an IELTS score of 5.5.

Table 32

Summary of Predictors of Posttest Performance in EIT for Visual Group

	Model 1	
Variable	**Est.**	**SE**
Constant	- 0.27	(0.68)
Pretest	0.64***	(0.108)
WM	- 0.09	(0.11)
DS	0.08	(0.12)
IELTS (ref. 5.5)		
Score 6.0	0.39	(0.51)
Score 6.5	0.43	(0.70)
R^2	0.49	

Note. $n = 39$; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; *** $p < 0.001$.

4.5   Discussion

This experiment was conducted with L1 Chinese speakers to examine the extent to which exposure to L2 syntax in each modality (auditory or visual) can stimulate automatic processing in the L2 system, resulting in the successful acquisition of implicit grammatical knowledge (Research Question 1). The language development accrued from each modality intervention was assessed by the measures of implicit knowledge (i.e., the timed GJT and the EIT). The experiment further aimed to explore whether WM capacity, as measured by the OSPAN task, predicts how this automatic processing is acquired differently with exposure to one of the two modalities (auditory vs. visual, Research Question 2).

4.5.1   Modality

The significant increase in performance between pre- and posttests in each modality group, determined by both the GJT and the EIT, was interpreted as evidence that inputs in either modality intervention (auditory or visual) were able to boost the automaticity of syntactic processing. However, the results demonstrated no differentiable learning effects between the visual and auditory interventions in promoting this automaticity, although the difference was marginally significant in both the GJT and the EIT measures ($p = 0.052$ and $p = 0.08$, respectively). In particular, analysis of participants' accuracy in the GJT indicated a greater mean improvement in the visual intervention modality than in the auditory intervention modality. A similar trend was also observed in the analysis of participants' accuracy in the EIT. The two modality groups did not differ significantly at the pretest based on the GJT and the EIT scores, indicating that this apparent tendency towards a visual modality advantage in grammatical knowledge automatization was most likely due to the visual intervention itself and not due to pre-existing differences between the two groups.

The observed trend towards the visual modality advantage corroborates the general theoretical claims about the difference in language performance depending on the encoding

modality. Visual linguistic input is assumed to be segmented into units more neatly (i.e., words and sentences) than the auditory input streams (Anderson, 1985), and consequently, deeper processing of linguistic information is produced. This is because the orthographic representation of the visual input is unvarying; that is, the visual stream is presented in a pattern of text that does not vary from one author to another. This is in contrast to the processing of the auditory modality, where listeners must deal with prosodic and intonational cues that vary from speaker to speaker to find word boundaries within a sentence (Alderson, 2005b). As a result, heavy demands are placed on the processing resources, and this most probably has a detrimental effect on successful auditory comprehension.

Other possible factors that account for the tendency towards better acquisition in the visual modality were repetitive processing and control over the input rate, which the visually presented language allows during the input processing. These monitoring processes can reduce the incidence of false recall and recognition (Gallo, 2006). This is unlike the real-time processing inherent in the nature of auditory input processing, which lacks cognitive strategies that promote access to auditory language comprehension, such as backtracking, skimming, and paying attention to/disregarding certain features of the stream.

The results were, indeed, supported by previous empirical studies (e.g., Johnson, 1992; Murphy, 1997; Park, 2004; Wong, 2001) that examined the difference in language performance depending on the encoding modality. Murphy (1997) found that modality clearly interacts with grammaticality where L2 learners were less accurate and slower to judge ungrammatical sentences (using the untimed GJT) in the auditory than in the visual modality. Park (2004) further observed a significantly better performance on the reading comprehension than in the listening comprehension test. He claimed that the difference in comprehension performance is due largely to the control over the input rate and the nature of decoding a text inherent in visual processing. Wong (2001) found that attending to form and

meaning simultaneously in the auditory modality negatively affects comprehension but not in the case of visual modality. Wong's results could explain the observed results in the present study; as the nature of auditory modality involves the processing of prosodic elements in addition to linguistic elements, the attention to the linguistic form of the input competes for learners' limited attentional resources with attention to meaning, resulting in a reduction in the amount of resources available for the processing of comprehension.

Nevertheless, given that the difference in performance between the two modality interventions failed to reach statistical significance in both the GJT and the EIT, a firm conclusion cannot be drawn that exposure to the visual modality can result in more successful acquisition of implicit grammatical knowledge than with the auditory modality. Therefore, if the lack of a significant difference was robustly taken into account, the findings of this experiment would also be in accordance with the results reported in several previous studies, which failed to find any differential effects in favour of the visual input modality (e.g., Morgan-Short et al., 2018) or to reveal fairly contradictory findings (e.g., Sagarra & Abbuhl, 2013; Sydorenko, 2010).

Morgan-Short et al.'s (2018) multi-site study, for instance, where the protocol at each site was strictly controlled, investigated the role of modality (visual vs. auditory) on input processing depending on the allocation of attention (meaning vs. form). The results did not demonstrate that attending to meaning/form simultaneously had any effect on comprehension when input was delivered through an auditory or visual modality. However, only in one site (US) did the results of a timed written task reveal that attending to a lexical form appeared to reduce comprehension. Therefore, the authors claimed that if the study had been conducted only in the United States, the conclusions about the results would have been different; that is, attending to a lexical form would interfere with comprehension.

More importantly, the results are incompatible with those of Sagarra and Abbuhl (2013), and of Sydorenko (2010), who investigated the relative effects of modality on learning itself. In particular, their results showed counter evidence in some tests, and most of these tests revealed a 'match' case between the modality of the learning and the modality of the tests. Sydorenko's (2010) study, which examined the effects of input modality on the vocabulary learning, found that exposure to video without captions resulted in the higher auditory recognition of vocabulary than a video with only captions (i.e., the video lacked audio) while the reverse was true with the visual recognition of vocabulary (i.e., dual modality and only text captions obtained greater performance for visual than auditory recognition tests of vocabulary). Similarly, Sagarra and Abbuhl's (2013) study, which investigated the role of feedback modality on L2 linguistic accuracy, found orally enhanced feedback had additional effects compared to typographically enhanced feedback on the interactional posttests (but not on the written posttests). The match case observed in these studies did not appear in the current results. One possible explanation relates to the type of tests used in measuring language development. That is, the tests used in Sagarra and Abbuhl's (2013) and Sydorenko's (2010) studies are biased in favour of explicit rather than implicit knowledge. This is unlike the language tests used in the current investigation (i.e., the EIT and the timed GJT), where both tests are presumed to provide more critical data in terms of L2 acquisition (i.e., measuring implicit knowledge acquisition).

Drawing together the discussions above, it is possible to conclude that visual input is more effective than auditory input in promoting automaticity, leading to the successful acquisition of L2 grammatical knowledge. This is because the observed trend towards the visual modality advantage in language acquisition is in accordance with the general claims and many empirical results of studies that investigated modality effects on language performance rather than learning. The possible variable which made the present results

different from those of Sagarra and Abbuhl (2013) and Sydorenko (2010) is that language development in the present investigation was assessed using measures of implicit knowledge. Thus, Doughty (2004) stressed that the valid measure of the impact of any kind of learning is that it involves real deployment of L2 knowledge under any condition such as spontaneity (i.e., implicit knowledge).

4.5.2 Syntactic structure

Improvement in performance across the three structures — tag questions (TQ), negative adverbs (NA), and counterfactual conditional (CON) — in the auditory group is not significantly different from that observed in the visual group, suggesting that the greater improvement resulting from visual intervention was not influenced by the variation in performance on syntactic structures between the groups. However, across the two modality groups, a varied amount of learning across the three structures was noticeable in evidence in the EIT performance but not in the GJT, where the mean improvement in the TQ was significantly lower than that observed in the NA and CON structures. Similar results were present in the visual group. In the auditory group, however, the difference lay only between the TQ and NA structures.

To determine the variation in the amount of learning across the structures in the EIT more deeply, analysis of the data in performance on the EIT indicated that of the three structures used in stimuli, the TQ items elicited the highest mean score in the pretest performance. This result was also evident in the GJT, suggesting that participants apparently had fairly thorough implicit knowledge of the TQ structure prior to their exposure during treatment, compared to the NA and CON structures. TQs reflected a very small average increase in the score ($M = 0.10$). This small improvement may be explained by the high average score in the pretest, resulting in participants having difficulty in achieving a greater increase in performance on the posttest (i.e., ceiling effects). Moreover, from a

psycholinguistic perspective, the poor improvement probably occurred due to the two TQ-inherent factors influencing its conceptual complexity (i.e., complex function and difficult processing). Processing this structure is complicated due the requirement for procedures pertaining to the processing of a subordinate clause, based on Pienemann's (1999) hierarchy of processing procedures. In particular, processing of a TQ necessitates referring back to the verb phrase in the main clause to check agreement in tense and subject; as a result, greater processing demands are likely to be placed on automaticity when it is specifically presented as part of the EIT, with good performance entailing high fluency in auditory input and oral output, both of which hinge on automatized processing. Therefore, a lack of automaticity in either the auditory input processing or the oral output processing is sufficient to be detrimental to performance. In this respect, the EIT is unlike the timed GJT, which primarily requires only the fluent visual recognition of linguistic forms. Therefore, the same structure (TQ) elicited greater improvement in the GJT than that observed in the EIT.

Analysis of the NA data from the EIT indicated that this structure elicited a lower performance than for TQ in the pretest. However, of the three structures, the NA structure reflected the most noticeable improvement between the pre- and posttests. A possible explanation that might account for the greater improvement is that the NA structure's syntax does not involve a subordinate clause to which a participant needs to refer back to build a syntactically correct sentence. As a result, processing of the NA structure demanded less cognitive effort in the EIT than the processing of the TQ structure. This superior improvement was obtained although the syntactic rule governing the NA structure does not meet the regularity criterion: the presence of the NA necessitates inversion of the auxiliary verb with the subject. This inversion contrasts with other adverbs, such as those pertaining to time, place, and manner, which follow the normal verb order in English syntax. That is, the

auxiliary verb is positioned after the subject even when negation is embedded (e.g.,

Normally, my children don't watch TV in the evening).

The abnormality of this word order is expected to make the NA syntax burdensome to

become competently automatized because abundant cognitive effort is required for the L2

learner's grammatical knowledge to be re-automatized such that he or she can generate an

auxiliary verb-subject inversion without conscious consideration or with only limited

consideration. However, improvement in the performance on NA items might be promoted

by its requirement of one clause and its position being close to the beginning of a sentence

(e.g., *Seldom **had he** seen a fat football player*).

Analysis of EIT data pertaining to the CON structure reflected the poorest

performance on the pretest, perhaps showing that participants had very restricted implicit

knowledge of the CON structure. Although the structure clearly reflected less improvement

than that observed in the NA structure, the improvement reflected a distinct learning effect

compared with the TQ. The weak improvement pattern in the CON is most probably

explained by the three factors contributing to conceptual complexity, which are in operation

in the rule system governing the CON structure, specifically, its perceptual non-salience,

irregularity, and complex processing, all of which predict processing difficulties. The latter

(complex processing) represents increasing support for the assumption, previously suggested

in relation to the TQ structure, that a syntactic structure involving a subordinate clause places

high demands on linguistic processing resources when specifically presented as part of the

EIT due to the increased requirement for automaticity in this task format (i.e., it requires the

processing of auditory input and oral output). This is also supported by the mean

improvement in performance on the CON structure in each modality group. That is, the

improvement in the CON performance reflected a distinct learning effect from the TQ in the

visual group, but not in the auditory group (i.e., due largely to the distinctive idiosyncrasies inherent in auditory processing which had a detrimental effect on the CON performance).

4.5.3   Grammaticality

The tendency towards better learning in the visual modality did not interact with the type of grammaticality tested. That is, improvement in performance between grammatical and ungrammatical items in the auditory group is not significantly different from that observed in the visual group. Across the two modality groups, similar patterns of mean improvement were in evidence for both grammatical and ungrammatical items. That is, the participant was able to correctly reject ungrammatical items in the case of the GJT and generate oral fluent correction of ungrammatical sentences in the case of the EIT. Similar results were also in evidence in each individual group. This suggests that the degree of automated knowledge of L2 syntax was increased with either modality intervention without being influenced by the type of grammaticality.

However, the difference in performance between grammatical and ungrammatical items was observed in the visual group only when improvement in performance on the different structures were analysed based on grammaticality. The results revealed comparing improvement in performance on the CON structure to the TQ structure differed significantly between grammatical and ungrammatical items (i.e., performance on the CON structure was noticeably lower for ungrammatical items than that observed for grammatical items whereas the opposite was observed for the TQ; performance on the TQ structure was better for ungrammatical items than that observed for grammatical items, see Figure 15 and Figure 16).

These results might support the claim that responses to grammatical and ungrammatical sentences draw on two distinct cognitive processes. Ellis (1991, p. 178) observed that "sentences that learners judged to be ungrammatical or that they were not sure about often invoked attempts to make use of declarative knowledge". This is based on the

hypothesis that more cognitive effort needs to be exerted to process an ungrammatical item, and consequently, explicit knowledge pertaining to the ungrammatical element will be retrieved (more discussion was provided in Experiment 2).

To summarise the discussions pertaining to the first research question, the results provided evidence that inputs in either modality intervention (auditory or visual) can effectively promote the automaticity of syntactic processing as measured by both the timed GJT and the EIT. Although the results demonstrated no differentiable learning effects between visual and auditory interventions, a better performance was observed in the visual modality in both the GJT and the EIT, suggesting an apparent tendency for the process of automatizing L2 grammatical knowledge to be fostered more successfully via the visual modality than via the auditory modality. The tendency towards better acquisition in the visual modality was not linked with the type of grammaticality or syntactic structures tested. Across the whole sample, the variation in the amount of acquisition across the three structures was observed only in the EIT, where the mean improvement in the TQ structure was significantly lower than that observed in the NA and CON structures. Similar results were present in the visual group. In the auditory group, however, the difference lay only between the TQ and the NA structures. Across the whole sample and in each modality group, similar patterns of improvement were also found with both grammatical and ungrammatical items. This suggests that the knowledge of L2 syntax was automatized to such an extent that ungrammatical items were correctly responded in both the GJT and the EIT. Only in the visual group, improvement on the different structures was influenced by grammaticality.

4.5.4 Working memory

The results demonstrated that WM capacity plays a distinct role in enhancing L2 automaticity for auditory versus visual modality-based language learning. However, this effect was apparent only in a certain modality and for particular item classes. More

specifically, WM capacity, as measured by the OSPAN task, was found to correlate

positively with overall improvement in performance on the GJT for the auditory group but

not in the EIT. Further analysis of the GJT data demonstrated that this observed correlation

was driven by the correlation between the OSPAN task and performance on ungrammatical

items and items of the NA structure. One possible reason resulting in the lack of the

relationship between WM capacity and the EIT scores in the case of auditory training may

pertain to poor performance on the EIT itself. The EIT elicited a poorer average improvement

score from pre- to posttest ($M = 2.18$, $SD = 0.50$) than the timed GJT ($M = 3.18$, $SD = 0.61$).

This small improvement in the EIT might trigger a small variation between the high and low

automated knowledge of L2 syntax, resulting in a difficulty in distinguishing individuals with

high and low WM capacity in the efficiency of EIT processing (see Section 6.4 for more

discussion about the difference between the timed GJT and the EIT performance). This

interpretation can be supported by the lack of relationship between phonological short-term

memory (as measured by a digit span task) and the EIT performance.

In contrast, no significant correlation was observed between WM capacity and the

visual exposure condition in the overall accuracy in both measures, namely, the GJT and the

EIT. Taken together, the obtained results on the relationship between WM capacity and the

timed GJT suggest that higher WM capacity provides a greater benefit to acquisition in the

auditory than the visual modality. This is distinct from the results relating to visual input;

individuals with high and low WM capacity do not differ significantly in the efficiency of

their visual input processing.

The distinction in the encoding modality is mostly manifested in access to language

input: specifically, access to L2 auditory input is more restricted than access to visual input.

To illustrate this more clearly, consider the NA structure in a sentence like *Seldom do we get*

the chance to watch television together as an example. The present results revealed that WM

capacity predicts the successful acquisition of NA items in only the auditory exposure intervention. In this case, the processor must parse the auditory stream by recognising acoustic elements like intonation, stress pattern, and accentual system, and further linguistic elements like phonemes, syntax, lexical items, the beginnings and endings of words, etc. Furthermore, the L2 listener must associate the utterance heard with his or her background and contextual knowledge in order to construct a conceptual framework for meaningful linguistic comprehension.

The challenge lies not only in executing all of the above-mentioned processes, but also in doing so in real time; in turn, it may prove challenging for L2 learners to unpack linguistic meaning effortlessly from the incoming input. Put differently, in the case of low WM capacity, on encountering the string *Seldom do we get*, the parser is most likely to encounter difficulty in processing this abnormal word order (i.e., the inversion of the subject and auxiliary) if the listener's linguistic knowledge of the NA structure has not been automatized. This implies that more factors, including conscious attention, compete for the learner's limited WM resources during the processing of this initial string, and greater cognitive effort is most likely to be expended specifically on the syntactic abnormality (i.e., more controlled processing is carried out). As a result, the amount of WM resources devoted to executing this controlled processing reduces the resources available to engage in other processing tasks (i.e., lexicon, semantics, propositional content, etc). That is, participants with a low WM capacity in the auditory intervention appeared to struggle to parse the rapid stream of the NA sentence into meaningful constituent parts, and consequently, their capacity for significant learning was seriously impaired. On the contrary, participants with a high WM capacity were able to execute the parallel cognitive processes involved in processing an auditory L2 stream more efficiently and rapidly than those with a low WM capacity.

In the case of visual exposure, a participant with a low WM capacity, on encountering the same string *Seldom do we get the chance to watch television together*, might not struggle greatly to parse it as is the case with auditory exposure. This is because the elements associated with prosody are absent; that is, the orthographic representation of the NA sentence is mutable. Secondly, the visually presented language allows for close, repetitive processing, and thus the processor is able to control the rate of visual input. These facilitative factors inherent in the nature of visual exposure, indeed, increase the possibility of multiple processes functioning simultaneously. As a result, the demands exerted by the initial string of the sentence *Seldom do we get* on WM capacity are minimal because fewer factors compete for WM resources during the processing of the visually presented NA sentence (i.e., distractibility is lessened). Consequently, an adequate amount of WM resources can be allocated to processing the initial syntactic parse without major detrimental effects on the amount of resources available to engage in other processing tasks.

The obtained results are broadly consistent with the general theoretical claims suggesting that WM mediates the effects of auditory input on language performance. The results are partly in line with previous empirical studies on the relationship between L2 auditory comprehension (including learning) and WM (e.g., Andringa et al., 2012; Brunfaut & Révész, 2015; Kormos & Sáfár, 2008; Vandergrifta & Baker, 2015). This might be due to the dearth of previous studies addressing this relationship, leading to inconclusive evidence. By contrast, the obtained results are broadly inconsistent with the major trends of previous studies investigating the relationship between WM and L2 visual comprehension (e.g., Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Leeser, 2007; Walter, 2004).

Kormos and Sáfár's (2008) study, for instance, which investigated the relationship between WM capacity and L2 performance in several language skills, found that both L2 auditory and visual comprehension were predicted by WM capacity, with a slightly higher

correlation with the auditory ($r = 0.37$) than the visual ($r = 0.31$) comprehension. Other studies (e.g., Andringa et al., 2012; Vandergrifta & Baker, 2015) suggested that WM is a weak predictor of auditory comprehension when other factors are taken into consideration. In Vandergrifta and Baker's (2015) study, for instance, the findings revealed that WM tends to be related to L2 auditory comprehension, where a significant relationship was obtained only for the first group, but it was not sufficiently strong to elicit significant results with the second and third groups (participants were sampled based on the calendar year they joined the program: 2008, 2009 and 2010). Furthermore, the results are also in partial agreement with what Brunfaut and Révész (2015) found in their study in that moderate correlations were observed between WM capacity and the overall listening performance section of the Pearson Test of English, which is designed to measure L2 listeners' local comprehension. However, the researchers did not find significant correlations with the listening test designed to measure L2 listeners' global comprehension. The authors concluded that a listening task that is designed to emphasise global comprehension might tax WM resources less than a local comprehension-based listening task involving deciphering and maintaining deep information.

More importantly, the results parallel evidence from Sagarra and Abbuhl (2013), who found WM capacity is correlated with the amount of learning accrued from auditory feedback, but not from visual feedback, based on language tests (written and oral). Although Sagarra and Abbuhl (2013) did not use measures of implicit language knowledge, these measures were employed in the current investigation. One could argue, thus, that the correlations elicited by the analysis, indeed, result from the difference in processing demands exerted by modality on WM capacity.

In relation to the relationship between grammaticality and WM capacity, the results revealed that WM predicted successful acquisition for improvement in performance on ungrammatical items of the timed GJT only in the case of auditory modality. A syntactically

deviating element in sentences is most likely to put pressure on the parser by imposing computational complexity, and in turn, WM capacity is taxed to a greater extent. As elucidated above, in the case of low WM capacity, more cognitive factors compete for the learner's limited WM resources over the processing of the deviating element in a sentence; consequently, the capacity for significant learning is seriously impaired. On the contrary, participants with a high WM capacity are able to achieve parallel cognitive processes to process syntactically deviating elements.

The results pertaining to WM-grammaticality are inconsistent with what was found in Tagarelli, Mota and Rebuschat's (2015) study, in which WM capacity was revealed to predict performance only in grammatical items in an explicit learning condition but not in an implicit learning condition. Although the authors concluded that "the differential effect of WM capacity on grammatical and ungrammatical items suggests that this relationship is complex" (p. 242), the present investigation was able to provide evidence to support the general claim, suggesting that ambiguous sentences (e.g., ungrammatical structures) are assumed to tax WM capacity (Juffs, 2015). One possible explanation which accounts for this obtained result is that it is due to the type of language task with which WM capacity correlates. Put differently, WM capacity in the present investigation correlated with the timed GJT, which is assumed to measure implicit knowledge, while WM capacity in Tagarelli, Mota and Rebuschat's (2015) study was found to correlate with the untimed GJT, which is assumed to measure explicit knowledge. The timed GJT is assumed to measure the level of automaticity in the participant's acquired grammatical knowledge. In particular, if the knowledge of L2 syntax was automated to such an extent that ungrammatical items were correctly responded to, then the computational demands on WM capacity are minimal. However, if the knowledge of L2 syntax was not automated to the extent whereby ungrammatical items were correctly identified, then computational demands would impose on WM capacity, leading to a

difference between L2 learners with high and low WM capacity regarding the efficiency of their judgments in processing.

To summarise the discussions relating to the second research question, the results revealed that WM interacted only with auditory input-based language learning, as determined only by the timed GJT (a measure of implicit knowledge). These critical data lend further credence to the hypothesis that although auditory and visual comprehension depend on identical basic processes (i.e., the interaction of top-down and bottom-up processing), the processing demands exerted by auditory linguistic input on WM capacity are greater than those exerted by visual input, leading the former to distinguish individuals with high and low WM capacity due to the efficiency of their auditory input processing. The results further indicated that WM capacity was found to predict successful acquisition for improvement in performance on items of NA and ungrammatical items. Success in obtaining a WM-ungrammatical items interaction, unlike Tagarelli, Mota and Rebuschat's (2015) study, is possibly explained by the type of measures employed to assess language development. That is, measures of implicit knowledge are presumed to provide more critical data and elucidate whether the WM capacity is susceptible to computational complexity (i.e., increased processing demands) via auditory or visual input-based language learning. The elicited data related to the first and second research questions provided relatively cohesive results; that is, as the visual intervention tended to be more effective than the auditory intervention at fostering L2 automatic processing, the auditory intervention, as determined by timed GJT, was found to correlate with WM capacity.

Chapter 5: Experiment 2

5.1   Introduction

Experiment 2, like the Experiment 1, aimed to examine whether there is a significant difference in the relative effects of auditory and visual input modalities on the acquisition of implicit grammatical knowledge. The experiment further aimed to determine whether WM capacity mediates the effects of auditory or visual input modality on the acquisition of implicit grammatical knowledge. Unlike Experiment 1 which targeted L1 Chinese language, participants in Experiment 2 were L1 Arabic speakers.

5.2   Participants

A total of 37 L1 Arabic speaker participants took part in this study. They were randomly assigned to one of two groups: auditory group (10 male and 8 female, aged 18–36 years '$M = 26$') and visual group (10 male and 9 female, aged 22–43 years '$M = 30$'). Most of them were students who were either studying English or who were a part of an undergraduate or postgraduate degree programme in Bristol and nearby cities. They were intermediate-level speakers of English. This English level is the equivalent of a score of 5.5–6.5 on the academic version of the International English Testing System (IELTS) or equivalent scores on the Test of English as a Foreign Language (TOEFL, only one participant had a TOFEL score), provided that it was a last score that a participant received. The period between the last IELTS score and the time when the participants took part in the experiment was not strictly managed in order to maximise the opportunity of recruiting as many potential participants as possible. This means that most participants took the IELTS within the last 2 years at the most, except 13 % of participants whose IELTS scores were gained more than two years before.

As noted earlier, the pretest performance in the timed GJT was the main measure which was used to determine the extent of the participants' prior knowledge of the stimulus material. Accordingly, no participants were eliminated from the analysis as they obtained above 19 out of 24 in total. However, three participants were removed from the analysis because they did not attend the second session, did not perform the whole second session, or failed to carefully read the tasks' instructions. Participants received £14 for participating in the experiment.

5.3   Arabic syntax

Arabic is an ancient Semitic language that belongs to the Afro-Asiatic linguistic family. Semitic languages include extinct members such as Phoenician, endangered languages such as Aramaic, and survivors such as Hebrew and Arabic (Swan & Smith, 2001). Arabic is the native language of over 200 million people in 20 independent Middle Eastern and African countries such as Oman, Egypt, Sudan, the United Arab Emirates, Saudi Arabia, Lebanon (Ryding, 2005).

Because Arabic is the language of the Islamic religion, the existing written Arabic has been mostly unchanged. This is due largely to its position in Islam. McLoughlin (2002, p. 1) claimed, "It is the Quran which has preserved the essence of written Arabic, and it is also the elevated status accorded to the original language of Islam which has prevented the Arabic dialects from becoming as far apart from each other as the dialects of Latin". This accounts for the fact that the citizens of Arab countries are still able to communicate with each other despite the geographic remoteness between their countries.

Arabic has two varieties: standard and colloquial. Standard Arabic, as described by Holes (2004, p. 5), is "the modern descendant of Classic Arabic, unchanged in the essentials of its syntax, but very much changed, and still changing, in its vocabulary and phraseology".

Standard Arabicis also the language of literacy and the official language which contacts all Arab countries. This is manifested in news, political speeches, and textbooks in education (Ryding, 2005).

However, colloquial Arabic is the Arabic dialect that native speakers acquire during their childhood before they begin formal school, based on the geographic area from which they come. It is the language of the streets and everyday life (Ryding, 2005). Speakers from any region or town normally do not find it difficult to communicate with speakers of the vernacular dialect of the neighbouring town or region. More specifically, the greater the distance between the spoken tongues of Arab nations, the greater the difference between the linguistic form in vernacular speech. The difference is often greater than the differences between English dialects, such as the differences between UK, US, Canadian, and Australian English. Thus, it is easy to detect that Arabic dialects used in extreme areas such as Morocco in the west are substantially distinct from those spoken in the other extreme areas such as Oman in the east. Even small regions within a country can have different vocabularies and, in some situations, even distinct syntactic forms (Swan & Smith, 2001).

Whereas English has a fixed word order: subject-verb-object (SVO), Arabic possesses a comparatively free word order. This means multiple word order permutations are permitted (SVO and VSO), as well as distinct inflectional patterns. In the SVO word order, the verb must agree in person, gender, and number with a full NP subject, whereas in the VSO word order, the verb must agree with its subject only in person and gender (Ayoun, 2005). Al-Momani (2010, as cited in Abu-Shquier & Abu Shqeer, 2012, p. 51) claimed that the word order pattern in Arabic is selected based on prior information. That is, if a subject has not been already introduced, the VSO order is favoured. On the contrary, if a subject has been already introduced, the SVO order is most suited for use. The following are examples of VSO and SVO word orders.

Table 33

An Illustration of a VSO and an SVO Word Order in Arabic

	Word order			Word order		
	V	S	O	S	V	O
Sentence	يكتب	الطالب	الرسالة	الطالب	يكتب	الرسالة
Transliteration	*Yaktub*	*Attaleb*	*Arresalah*	*Attaleb*	*Yaktub*	*arresalah*
Gloss	Writes	the student	the letter	the student	Writes	the letter
Translation	the student	Writes	the letter	Writes	the student	the letter

5.4 Relevant syntactic structures in Arabic

5.4.1 Tag question (TQ)

Unlike English, Arabic has a fixed form of the tag question (TQ), which is interpreted as *isn't it so?* This means that the same tag precedes both positive and negative sentences and is usually marked by a rising pitch pattern in which the speaker demands confirmation of or agreement with what he or she is stating (Kebbe & Kebbe, 2000). The TQ in Arabic does not require agreement with the subject of the sentence in terms of number and gender. This suggests that the tag is introduced at the end of the sentence as an idiom with a particular semantic effect (i.e., as a request for confirmation or agreement, see Kebbe & Kebbe, 2000, for a review). The example below illustrates the difference in TQs between Arabic and English.

Arabic sentence: الطالبة لم تحضر الامتحان، أليس كذلك؟

Gloss: The student not attended the exam, isn't it so?

Translation: The student did not attend the exam, did she?

5.4.2 Conditional sentence (CON)

There is no one to one correspondence in 'if' particle between English and Arabic due to the difference in the syntactic system of conditionals between the two languages. However, in both languages, a *conditional sentence* consists of a dependent clause expressing one

situation (the *condition, antecedent,* or *protasis*) as a condition for the occurrence of another situation in the independent clause (the *result, consequent,* or *apodosis*). The three types of conditional sentences in English (factual, hypothetical, and counterfactual conditionals) are mostly introduced by the subordinating conjunction *if*. In Arabic, the type of conditional is determined by the use of various conditional conjunctions. For example, the most common conjunction used to introduce real conditional sentences is إذا *Etha* 'if'. Another conjunction used to introduce a real conditional sentence, which is used less frequently than إذا, is إن *En* (meaning: if in English). For the unreal conditional sentences in Arabic, there is a specific conjunction for the introduction of the dependent clause, which is لو *Law*. This is unlike English, where there is no specific conjunction for introducing unreal conditional sentences. In addition, the conjunction لو *Law* is most commonly used to express the past tense in unreal conditional sentences (see Schulz, Krahl, & Reuschel, 2000, for more details). The example below illustrates the difference in the past counterfactual condition between Arabic and English:

Arabic sentence: لو درست بجدٍ، لنجحت في الامتحان

Gloss: If you studied hard, you would pass the exam

Translation: If you have studied hard, you would have passed the exam.

5.4.3   Negative adverbs

Similar to English, Arabic language has negative adverbs, but these adverbs are limited compared to those in English, which possesses a variety of lexical words expressing negative adverbs. However, negative adverbs in Arabic language do not provoke the subject-verb inversion when those adverbs occur at the beginning of sentences as is the case in English. The following example illustrates the difference of negative adverbs between the two languages:

Arabic Sentence: نادراً جوليا تخرج ليلاً

Gloss: Rarely Julia goes out at night.

Translation: Rarely does Julia go out at night.

5.4.4 L1 transfer

Similar to the concept of L1 transfer in Experiment 1, comparing the differences and similarities between English and Arabic in terms of the stimuli structures could predict which are difficult and easy to learn. First, the basic word order of English (SVO) behaves in a similar way to Arabic because the word order of the latter can be SVO or VSO. This means that Arabic learners of English are likely to not find it difficult to learn the word order generally. In terms of stimuli structures, in particular, The comparisons revealed that the English tag question (TQ) does not behave similarly to its Arabic counterpart because the latter has only a fixed form of the question tag (*isn't it so?)* that applies to all grammatical and pragmatic situations in which the same tag precedes both positive and negative statements. Therefore, the level of difference in TQ structure between the two languages is low because Arabic learners learn the TQ almost like a new structure. Accordingly, the TQ is likely to be less difficult to learn for Arabic learners of English.

In the case of CON, the comparisons showed that Arabic has a counterfactual conditional, as is the case in English, but the dependent clause in Arabic must be introduced with a different conjunction from the factual conditional. The marked difference in CON between Arabic and English lies in the lack of present perfect in Arabic, which is required in the English main clause *if* and similar structure in an independent clause. Therefore, Arabic learners of English are likely to experience difficulty in learning the English CON structure. For the NA structure, although Arabic has a number of lexical words expressing negative adverbs, these adverbs do not cause subject–verb inversion when the adverb occurs at the

beginning of a sentence, as is the case in English. Because the point tested is the inversion in the NA structure, such structure is likely to be difficult to learn. Overall, it appears that all generated structures used in the current investigation are likely to be difficult for Arabic learners of English to learn due to L1 interference.

Table 34

Summarising the Level of Difficulty based on an Arabic–English Comparison for the Three Structures

Rule	Level of difficulty
TQ	Low
CON	High
NA	High

Note. Due to the L1 interference is not the focus of the current investigation, only basic comparisons were aimed at; TQ = tag questions; NA = negative adverb; CON = conditional.

5.5   Results

The results of Experiment 2 provide a detailed account of the statistical analyses carried out on the data obtained from Arabic learners of English to test the same research questions put forward in the experiment 1. That is, the analysis aims to investigate whether there is a significant difference in the EIT / timed GJT scores as a function of exposure to auditory input vs. visual input during language learning. This type of analysis allows for demonstrating which of two input modalities (auditory or visual) would elicit more successful acquisition of implicit grammatical knowledge (Research Question 1). The second aim of analysis was to explore whether there is a significant correlation between OSPAN task scores and the increase in performance (from pre- to posttests) on the EIT / timed GJT within each modality group. This type of analysis allows for determining whether WM capacity plays a

differential role on success in the acquisition of implicit grammatical knowledge under

exposure to the auditory and visual modalities (Research Question 2).

5.5.1   Analysis of timed GJT results

Table 35 shows participants' mean raw scores, with standard errors in both the pre-

and posttests in each of the two modality-specific interventions, including overall

performance and means at each level of the syntactic construction and grammaticality

conditions.

Table 35

Correct Mean (M) Raw GJT Scores with Standard Errors (SEs)

Condition	Pretest		Posttest		Improvement	
	M	*SE*	*M*	*SE*	*M*	*SE*
Overall performance	9.81	0.46	12.48	0.55	2.67	0.52
Auditory	10.00	0.57	13.38	0.80	3.48	0.79
Visual	9.63	0.72	11.63	0.74	2.00	0.67
Syntactic structure						
TQ	3.72	0.25	4.67	0.27	0.94	0.34
Auditory	4.05	0.35	5.22	0.37	1.16	0.53
Visual	3.42	0.36	4.15	0.36	0.73	0.43
NA	3.10	0.23	3.59	0.23	0.48	0.23
Auditory	3.05	0.34	3.72	0.32	0.66	0.39
Visual	3.15	0.31	3.47	0.33	0.31	0.27
CON	2.97	0.24	4.21	0.27	1.24	0.29
Auditory	2.88	0.34	4.44	0.42	1.55	0.44
Visual	3.05	0.37	4.00	0.35	0.94	0.38
Grammaticality						
GR	7.13	0.39	9.00	0.35	1.86	0.36
Auditory	7.50	0.42	9.55	0.37	2.05	0.52
Visual	6.78	0.65	8.47	0.57	1.68	0.51
UNG	2.67	0.24	3.48	0.34	0.81	0.35
Auditory	2.50	0.33	3.83	0.57	1.33	0.55
Visual	2.84	0.36	3.15	0.39	0.31	0.43

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items. The maximum raw score is 24 for overall performance, 8 for each syntactic structure, and 12 for each type of grammaticality (i.e., each type of syntactic structure includes four grammatical items and four ungrammatical items, for a total of 24).

Before carrying out inferential analyses to address the research questions, an independent sample t-test was performed on the pretest scores to rule out differences between the auditory and visual groups prior to the intervention. The analysis demonstrated no significant difference in GJT scores for auditory modality ($M = 10.00$, $SE = 0.57$) and visual

modality (M = 9.63, SE = 0.72) conditions, t (35) = 0.39, p = 0.69, indicating that the sample

was homogeneous and that any differences observed in the posttests could be attributable to

the modality intervention. Furthermore, an independent samples t-test was conducted on the

pretest scores to rule out differences between Experiment 1 and 2 prior to the intervention.

The analysis indicated a significant difference in the GJT scores for Experiment 1 (M =

12.49, SE = 0.37) and Experiment 2 (M = 9.81, SE = 0.46), t (112) = 4.30, p < 0.001,

suggesting that any differences in results between Experiment 1 and 2 might be due to

differences in participants' available implicit knowledge of their English syntax prior to

exposure to treatment interventions (i.e., developmental readiness hypothesis, see Section 7.2

for more discussion).

In addition, it is also important to check the data of each variable used in the

ANOVAs below in terms of the normality of distribution and outliers. The results of the

Shapiro-Wilks tests revealed that the data of the variables that pertained to answering the first

research question were normally distributed (i.e., the pre- and posttest performances for each

group were checked individually). However, two outliers were observed in the posttest for

the visual group. The following tables: Table 36, Table 37, Table 38, Table 39, and Table 40

show a summary of the data of all variables in terms of normality of distribution and outliers.

Table 36

Summary of Shapiro-Wilk Tests for GJT Data Across Both Modality Groups and in Each
Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Overall performance				
Pretest	0	37	0.95	0.09
Posttest	0	37	0.97	0.51
Modality				
Auditory				
Pretest	0	18	0.92	0.15
Posttest	0	18	0.92	0.17
Visual				
Pretest	0	19	0.92	0.15
Posttest	2	19	0.96	0.60

Table 37

Summary of Shapiro-Wilk Tests for GJT Data on Items of syntactic Structures Across
Groups and in Each Individual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Syntactic Structure				
TQ				
Pretest	0	37	0.95	0.103
Posttest	0	37	0.96	0.208
NA				
Pretest	0	37	0.95	0.104
Posttest	0	37	0.94	0.46
CON				
Pretest	0	37	0.95	0.108
Posttest	0	37	0.95	0.159

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

Table 38

Summary of Shapiro-Wilk Tests for GJT Data on Items of Grammaticality Across Groups and in Each Individual Group

Variable		Shapiro-Wilk Test		
		df	*F*	*P*
Grammaticality				
GR				
Pretest	1	37	0.95	0.13
Posttest	1	37	0.93	0.03
UNG				
Pretest	0	37	0.91	0.009
Posttest	2	37	0.94	0.09

Note. GR = grammatical items; UNG = ungrammatical items.

Table 39

Summary of Shapiro-Wilk Tests for GJT Data on Items of Syntactic Structures for Each
Type of Grammaticality in Auditory Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	0	18	0.81	0.002
Posttest	0	18	0.75	< 0.001
TQ × UNG				
Pretest	0	18	0.83	0.005
Posttest	0	18	0.89	0.055
NA × GR				
Pretest	0	18	0.91	0.087
Posttest	0	18	0.85	0.008
NA × UNG				
Pretest	1	18	0.79	0.001
Posttest	1	18	0.80	0.002
CON × GR				
Pretest	1	18	0.91	0.099
Posttest	1	18	0.73	< 0.001
CON × UNG				
Pretest	0	18	0.78	0.001
Posttest	0	18	0.83	0.005

Note. TQ = tag questions; NA = negative adverb; CON =
conditional; GR = grammatical items; UNG = ungrammatical
items; each syntactic structure: 8 items, and each type of
grammaticality: 12 items (i.e., each type of syntactic structure
includes 4 grammatical and 4 ungrammatical items, 24 in
total).

Table 40

Summary of Shapiro-Wilk Tests for GJT Data on Items of Syntactic Structures for Each Type of Grammaticality in Visual Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	0	19	0.90	0.05
Posttest	0	19	0.83	0.004
TQ × UNG				
Pretest	0	19	0.85	0.008
Posttest	0	19	0.87	0.016
NA × GR				
Pretest	0	19	0.90	0.069
Posttest	0	19	0.86	0.011
NA × UNG				
Pretest	1	19	0.77	0.001
Posttest	0	19	0.78	0.001
CON × GR				
Pretest	0	19	0.87	0.018
Posttest	0	19	0.76	< 0.001
CON × UNG				
Pretest	1	19	0.82	0.003
Posttest	0	19	0.84	0.007

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items; each syntactic structure: 8 items, and each type of grammaticality: 12 items (i.e., each type of syntactic structure includes 4 grammatical and 4 ungrammatical items, 24 in total).

As the descriptive statistics in Table 35 illustrate, the overall pattern across all conditions indicated improvement, as reflected in the higher mean scores obtained in posttests compared to the pretests. Across the sample as a whole, the mixed ANOVA showed a significant main effect (i.e., collapsing over all other variables) for the testing session, $F(1, 35) = 27.13$, $p < 0.001$, with a large effect size, partial eta-squared, $\eta p^2 = 0.43$ (i.e., based on

Cohen's scale, 1988, 0.01 = small effect, 0.06 = moderate effect, 0.14 = large effect),

reflecting an overall increase in the mean GJT score from the pretest (M = 9.81, SE = 0.46) to

the posttest (M = 12.48, SE = 0.55).

The overall pattern of the pre- and posttest data relating to different syntactic

structures revealed a significant main effect of syntactic structure, $F(2, 70)$ = 6.42, p = 0.003,

with a large effect size, ηp^2 = 0.15, demonstrating that there was a difference in mean GJT

scores across the three types of structure (TQ, NA and CON). Post hoc analysis with

Bonferroni adjustment showed that the average scores were significantly higher in the TQ

structure (M = 4.21, SE = 0.19) than in either the NA (M = 4.21, SE = 0.19) or the CON (M =

4.21, SE = 0.19) structures, p = 0.002, SE = 0.22, p = 0.03, SE = 0.22, respectively, whereas

the average scores for the NA (M = 3.35, SE = 0.20) and the CON structures (M = 3.59, SE =

0.21) were similar (p = 1.00, SE = 0.28), suggesting there was no difference in the mean GJT

scores between the two structures.

Similarly, the overall pattern of data pertaining to distinct type of grammaticality

indicated a significant main effect of grammaticality, $F(1, 35)$ = 186.38, p < 0.001, with a

large effect size, ηp^2 = 0.84, demonstrating that the overall performance in the two pre- and

posttests on grammatical items (M = 8.08, SE = 0.32) is significantly better than

ungrammatical items (M = 3.08, SE = 0.24).

The interaction between the testing session and syntactic structure was not statistically

significant, $F(2, 70)$ = 1.79, p = 0.19, ηp^2 = 0.04, demonstrating that the overall amount of

learning did not vary significantly for different structures: TQ, NA, and CON (M = 0.94, SE

= 0.34, M = 0.48, SE = 0.23, M = 1.24, SE = 0.29, respectively).

The analysis further indicated that the interaction between the testing session and

grammaticality was statistically significant, but with slightly below chance, $F(1, 35)$ = 4.36, p

= 0.044, with a large effect = 0.11, demonstrating that improvement from the pre-to posttests

in performance on grammatical items (*M* = 1.86, *SE* = 0.36) was significantly greater compared to ungrammatical items *M* = 0.81, *SE* = 0.35, as shown in Figure 18.
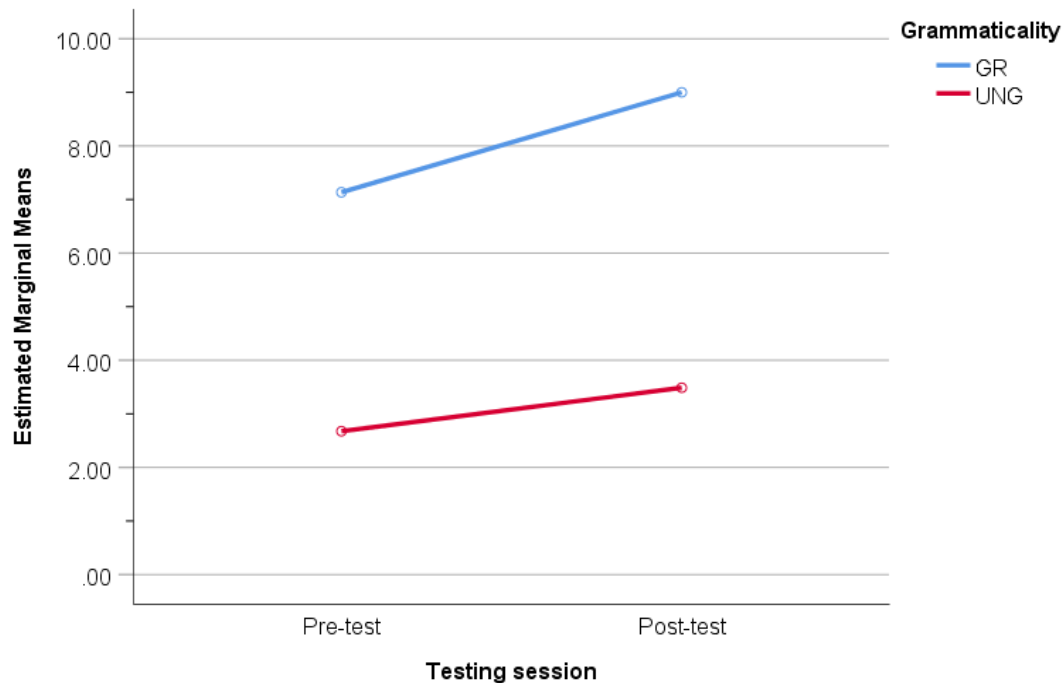

Figure 18. Performance for each grammaticality type in the pre- and posttest across both modalities.

However, when juxtaposing the two experiments regarding grammaticality x testing interaction, the results revealed no significant interaction between testing session, grammaticality and experiment, $F(1, 112) = 1.26$, $p = 0.26$, indicating that participants' improvement in performance on each of the grammaticality type was not influenced by the type of experiment (i.e.., L1 participants).

The descriptive statistics in Table 35 also display the pattern of data relating to modality interventions indicating a greater mean improvement from the pre- to posttests in the auditory intervention modality (*M* = 3.48, *SE* = 0.79) compared to the visual intervention modality (*M* = 2.00, *SE* = 0.67). This pattern was present in all individual conditions (i.e., both items of syntactic structures and grammaticality). However, the ANOVA analysis revealed no significant interaction between the testing session and modality, $F(1, 35) = 1.80$,

$p = 1.88$, $\eta p^2 = 0.04$, suggesting that performance in improvement did not significantly differ between the two modality interventions, as supported by the lack of overlap between the bars representing the auditory and visual interventions illustrated in Figure 19.
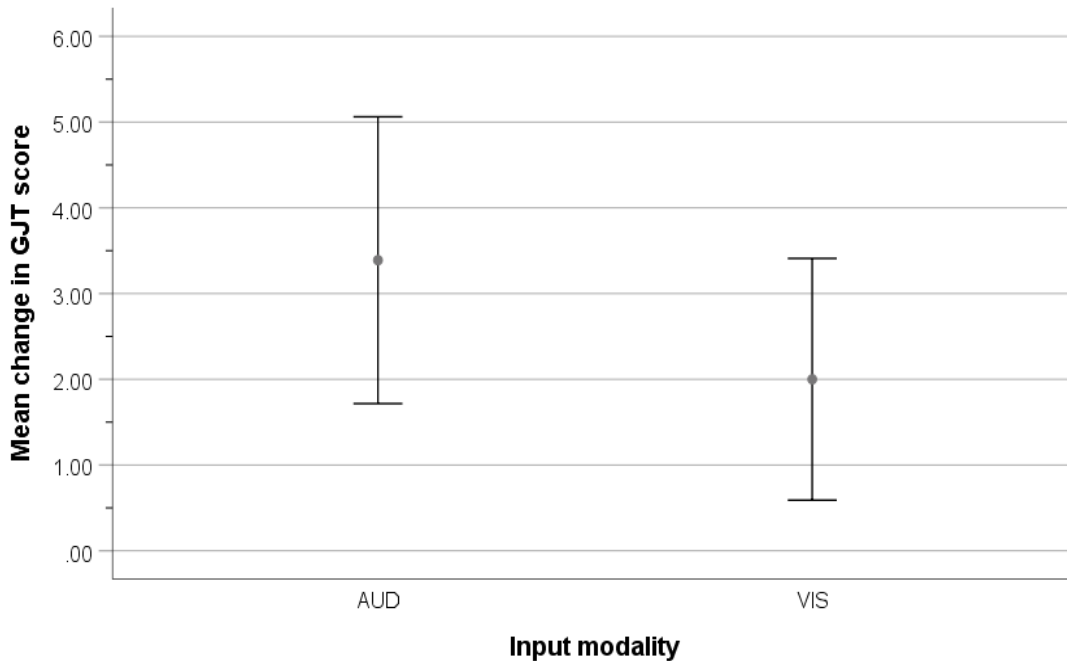
Figure 19. The difference in mean GJT improvement between auditory and visual modalities. Error bars indicate 95% confidence intervals.

As a differential improvement effect was not observed between the two modality interventions, there was no need for deeper analysis of relevant effects (e.g., to test whether any difference between the two interventions was influenced by type of syntactic structure or grammaticality).

Goal 2. The goal is to investigate whether auditory intervention participants' improvement in performance on each of the different syntactic constructions was influenced by grammaticality.

The results first indicated that there was a significant main effect of testing session, , $F(1, 17) = 18.27$, $p = 0.001$, with a large effect size, $\eta p^2 = 0.51$, demonstrating a significant improvement in the mean GJT score in the auditory group from the pre- ($M = 10.00$, $SE =$

0.57) to posttests ($M = 13.38$, $SE = 0.80$). However, the analysis showed no significant interaction between grammaticality and structure, $F(2, 34) = 0.68$, $p = 0.51$, $\eta p^2 = 0.03$, suggesting that grammatical and ungrammatical items did not influence participants' performance differently across the three structures. The analysis further revealed neither a significant interaction between the testing session and syntactic structure $F(2,34) = 0.91$, $p = 0.41$, $\eta p^2 = 0.05$, the nor testing session and grammaticality, $F(1, 17) = 0.98$, $p = 0.33$, $\eta p^2 = 0.05$. Furthermore, and more importantly, the ANOVA revealed no significant interaction between the testing session, grammaticality, and structure, $F(2, 34) = 1.00$, $p = 0.37$, $\eta p^2 = 0.05$, indicating that improvement in performance on each of the three different structures was not influenced by grammaticality.

Goal 3. The goal is to investigate whether visual intervention participants' improvement in performance on each of the different syntactic constructions was influenced by grammaticality.

In a similar vein, the results revealed there was a significant main effect of the testing session, $F(1, 18) = 8.88$, $p = 0.008$, with a large effect size, $\eta p^2 = 0.33$, indicating a positive change in the mean GJT score in the visual group from the pre- ($M = 9.36$, $SE = 0.72$) to posttests ($M = 11.63$, $SE = 0.74$). Moreover, the ANOVA analysis revealed there was significant interaction between grammaticality and syntactic structure, $F(2, 63) = 4.65$, $p = 0.01$, with a large effect size, $\eta p^2 = 0.20$, indicating that grammatical and ungrammatical items were associated with different effects on performance for different structures. The analysis further revealed neither a significant interaction between the testing session and syntactic structure $F(2,36) = 0.77$, $p = 0.46$, $\eta p^2 = 0.04$, nor testing session and grammaticality, $F(1, 18) = 3.94$, $p = 0.06$, $\eta p^2 = 0.18$. Furthermore, and more importantly, the ANOVA indicated no significant interaction between the testing session, grammaticality, and

structure, $F(2, 36) = 0.17$, $p = 0.84$, $\eta p^2 = 0.01$, indicating that improvement in performance on each of the three different structures was not influenced by grammaticality.

To summarise the results of analysis relating to the timed GJT, both auditory and visual exposure to L2 English syntax were found to result in successful acquisition of implicit grammatical knowledge. However, and more importantly, the results indicated no differentiable learning effects between visual and auditory interventions. For performance on syntactic structure across the two groups, no differing amounts of learning occurred for different syntactic structures, where similar results were present in each modality group. The results further showed that the improvement in performance on GR items were greater than UNG items. However, no significant difference was indicated in improvement between performance on grammatical and ungrammatical items within each modality group. Finally, participants' improvement in performance on each of the different structures was not influenced by grammaticality within each modality group (summary tables of the results are shown altogether with EIT's results under the Analysis of EIT Results section).

5.5.2 Analysis of EIT results

Table 41 shows participants' mean raw EIT scores, with standard errors, in both the pre- and posttests on the EIT for the two modality-specific interventions, including overall performance and means at each level of the grammaticality and syntactic structure conditions.

Table 41

Correct Mean (M) Raw EIT Scores with Standard Errors (SEs)

Condition	Pretest		Posttest		Improvement	
	M	*SE*	*M*	*SE*	*M*	*SE*
Overall performance	5.13	0.45	7.43	0.62	2.29	0.48
Auditory	5.16	0.60	7.88	0.74	2.72	0.75
Visual	5.10	0.68	7.00	1.00	1.89	0.61
Syntactic structure						
TQ	2.70	0.25	3.00	0.29	0.29	0.26
Auditory	2.66	0.30	3.50	0.38	0.83	0.30
Visual	2.73	0.40	2.52	0.42	-0.21	0.40
NA	1.81	0.19	2.70	0.24	0.89	0.29
Auditory	2.05	0.20	2.88	0.36	0.83	0.52
Visual	1.57	0.25	2.52	0.32	0.94	0.30
CON	0.62	0.17	1.72	0.32	1.10	0.26
Auditory	0.44	0.16	1.50	0.45	1.05	0.39
Visual	0.78	0.30	1.94	0.46	1.15	0.36
Grammaticality						
GR	4.29	0.39	5.97	0.49	1.67	0.40
Auditory	4.38	0.60	6.22	0.55	1.83	0.68
Visual	4.21	0.53	5.73	0.81	1.52	0.47
UNG	0.83	0.16	1.45	0.25	0.62	0.23
Auditory	0.77	0.22	1.66	0.34	0.88	0.33
Visual	0.89	0.25	1.26	0.37	0.36	0.34

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items. The maximum raw score is 24 for overall performance, 8 for each syntactic structure, and 12 for each type of grammaticality (i.e., each type of syntactic structure includes four grammatical items and four ungrammatical items, for a total of 24).

Before carrying out inferential analyses to address the research questions, a Mann-Whitney test was performed on the pretest scores to rule out differences between the auditory and visual groups prior to the intervention. The analysis demonstrated no significant difference in the EIT scores for auditory modality (*Mdn* = 4.5) and visual modality (*Mdn* = 5.0) conditions, U = 162.500, p = 0.79, indicating that the sample was homogeneous and that any differences observed in the posttests could be attributable to the modality intervention. Furthermore, an independent samples t-test was conducted on the pretest scores to rule out differences between Experiment 1 and 2 prior to the intervention. The analysis indicated no significant difference in the EIT scores for Experiment 1 (*M* = 6.18, *SE* = 0.33) and Experiment 2 (*M* = 5.14, *SE* = 0.45), but was close to being statistically significant, t (112) = 1.80, p = 0.07, suggesting that any differences in results between Experiment 1 and 2 might not be due to differences in participants' available implicit knowledge of their English syntax prior to exposure to treatment interventions (i.e., developmental readiness hypothesis see Section 7.2 for more discussion).

In addition, it is also important to check the data of each variable used in the ANOVAs below in terms of normality of distribution and outliers. The results revealed the data of the variables that pertained to answering the first research question were normally distributed except the pretest for auditory group. However, five outliers were observed: one in the pretest for auditory group and four in the pretest for visual group. The following tables: Table 42, Table 43, Table 44, Table 44, and Table 45 show a summary of the data of all variables in terms of normality of distribution and outliers.

Table 42

Summary of Shapiro-Wilk Tests for EIT Data Across Both Modality Groups and in Each Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Overall performance				
Pretest	5	37	0.94	0.07
Posttest	0	37	0.97	0.65
Modality				
Auditory				
Pretest	1	18	0.86	0.014
Posttest		18	0.96	0.74
Visual				
Pretest	4	19	0.93	0.20
Posttest	0	19	0.96	0.60

Table 43

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures Across Groups

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
Syntactic Structure				
TQ				
Pretest	0	37	0.95	0.140
Posttest	0	37	0.94	0.054
NA				
Pretest	0	37	0.90	0.005
Posttest	0	37	0.86	< 0.001
CON				
Pretest	3	37	0.65	< 0.001
Posttest	0	37	0.81	< 0.001

Note. TQ = tag questions; NA = negative adverb; CON = conditional.

Table 44

Summary of Shapiro-Wilk Tests for EIT Data on Items of Grammaticality Across Groups

Variable		Shapiro-Wilk Test		
		df	*F*	*P*
Grammaticality				
GR				
Pretest	1			
Posttest	1	37	0.97	0.45
UNG		37	0.95	0.16
Pretest	0			
Posttest	2	37	0.78	< 0.001

Note. GR = grammatical items; UNG = ungrammatical items.

Table 45

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures
for Each Type of Grammaticality in Auditory Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	0	18	0.92	0.18
Posttest	0	18	0.88	< 0.034
TQ × UNG				
Pretest	0	18	0.73	< 0.001
Posttest	0	18	0.79	0.001
NA × GR				
Pretest	1	18	0.92	0.155
Posttest	0	18	0.81	0.003
NA × UNG				
Pretest	3	18	0.45	< 0.001
Posttest	0	18	0.74	< 0.001
CON × GR				
Pretest	1	18	0.66	< 0.001
Posttest	0	18	0.80	0.002
CON × UNG				
Pretest	0	18	0	0
Posttest	3	18	0.43	< 0.001

Note. TQ = tag questions; NA = negative adverb; CON =
conditional; GR = grammatical items; UNG = ungrammatical
items; each syntactic structure: 8 items, and each type of
grammaticality: 12 items (i.e., each type of syntactic structure
includes 4 grammatical and 4 ungrammatical items, 24 in
total).

Table 46

Summary of Shapiro-Wilk Tests for EIT Data on Items of Syntactic Structures for Each Type of Grammaticality in Auditory Group

Variable	Outlier	Shapiro-Wilk Test		
		df	*F*	*P*
TQ × GR				
Pretest	0	19	0.90	0.052
Posttest	0	19	0.88	0.026
TQ × UNG				
Pretest	0	19	0.75	< 0.001
Posttest	0	19	0.71	< 0.001
NA × GR				
Pretest	1	19	0.90	0.072
Posttest	0	19	0.85	0.008
NA × UNG				
Pretest	1	19	0.24	< 0.001
Posttest	0	19	0.55	< 0.001
CON × GR				
Pretest	1	19	0.65	< 0.001
Posttest	0	19	0.82	< 0.003
CON × UNG				
Pretest	0	19	0.36	< 0.001
Posttest	4	19	0.52	< 0.001

Note. TQ = tag questions; NA = negative adverb; CON = conditional; GR = grammatical items; UNG = ungrammatical items; each syntactic structure: 8 items, and each type of grammaticality: 12 items (i.e., each type of syntactic structure includes 4 grammatical and 4 ungrammatical items, 24 in total).

As the descriptive statistics in Table 41 illustrate, the overall pattern across all conditions indicated improvement, as reflected in the higher mean scores obtained in posttests compared to pretests. Across the sample as a whole, the mixed ANOVA showed a significant main effect for the testing session, $F(1, 35) = 22.58$, $p < 0.001$, with a large effect

size, $\eta p^2 = 0.39$, reflecting an overall increase in the mean EIT score from the pretest ($M = 5.13$, $SE = 0.45$) to posttest ($M = 7.43$, $SE = 0.62$).

The overall pattern of the pre- and posttests data pertaining to the distinct syntactic structures indicated a significant main effect of syntactic structure, $F(2, 70) = 27.93$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.44$, demonstrating that there was a difference in mean EIT scores across the three types of structure. Post hoc analysis with Bonferroni adjustment revealed that the average scores were significantly higher in the TQ structure ($M = 2.85$, $SE = 0.24$) and in the NA structure ($M = 2.26$, $SE = 0.16$) than in the CON ($M = 1.17$, $SE = 0.22$) structures ($p < 0.001$, $SE = 0.20$, $p < 0.001$, $SE = 0.23$, respectively), whereas average scores for the TQ and the NA conditions did not significantly differ ($p = 0.06$, $SE = 0.24$).

Similarly, the overall pattern of data associating to the different types of grammaticality indicated a significant main effect of grammaticality, $F(1, 35) = 110.46$, $p < 0.001$, with a large effect size, $\eta p^2 = 0.75$, demonstrating that overall performance in the two pre- and posttests on grammatical items ($M = 5.14$, $SE = 0.40$) is significantly better than ungrammatical items ($M = 1.15$, $SE = 0.18$).

More importantly, the interaction between the testing session and syntactic structure was not statistically significant, $F(2, 70) = 2.27$, $p = 0.11$, with a moderate effect size, $\eta p^2 = 0.06$, indicating that the overall amount of learning did not vary considerably for different structures. These non-significant interactions are visualised in the interaction plot shown in Figure 20 below.
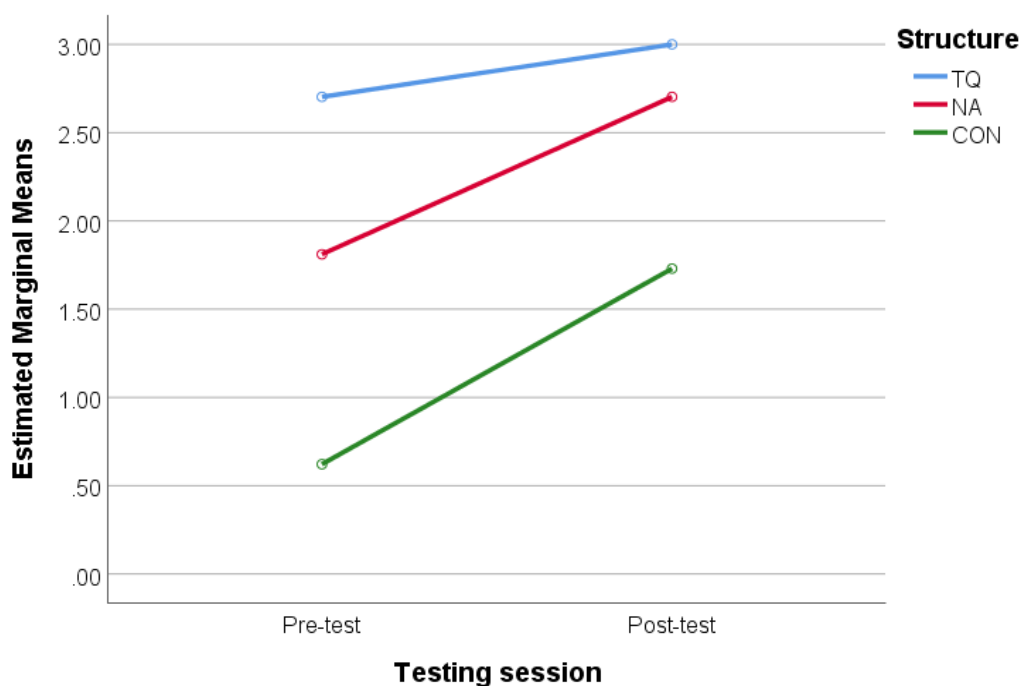
Figure 20. Performance foreach structure type in the pre-and posttest across both modality groups.

In a similar way, the analysis indicated that the interaction between the testing session and grammaticality was statistically significant, $F(1, 35) = 5.08$, $p = 0.03$, with a moderate effect size, $\eta p^2 = 0{,}12$ demonstrating that the overall improvement in performance on grammatical items ($M = 1.67$, $SE = 0.40$) was higher than improvement on ungrammatical items ($M = 0.62$, $SE = 0.23$), as supported by the differential slopes representing patterns of improvement for grammatical and ungrammatical items illustrated in Figure 21 below.
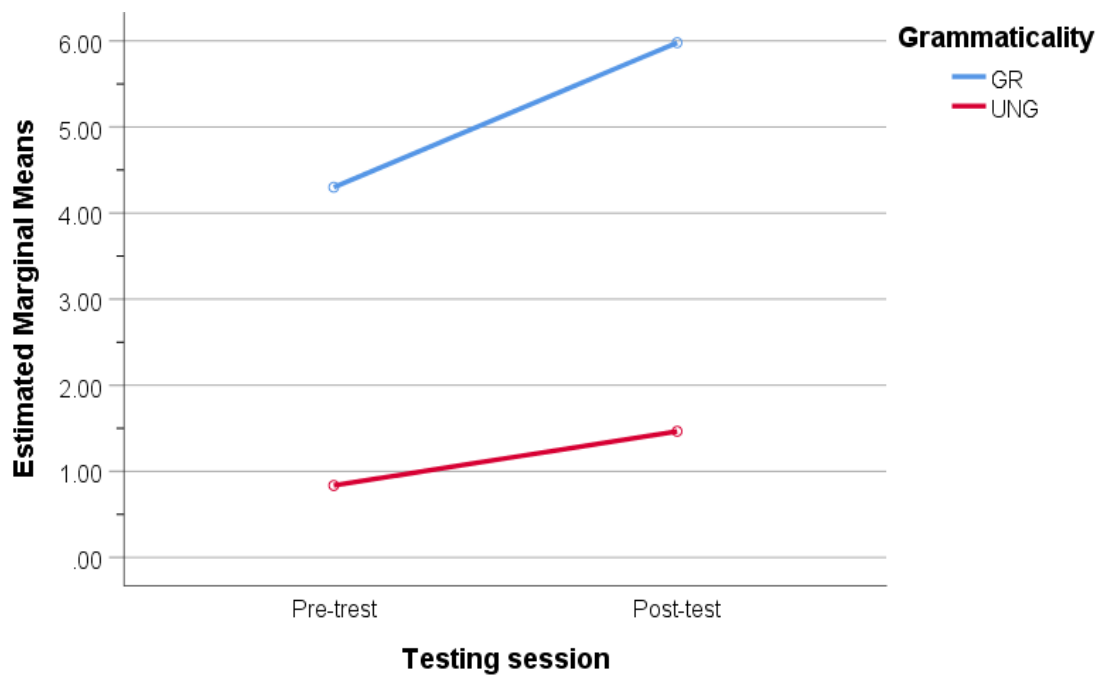
Figure 21. Performance for each type of grammaticality in the pre- and posttest across both modality groups.

However, when juxtaposing the two experiments regarding grammaticality x testing interaction, the results revealed no significant interaction between testing session, grammaticality and experiment, $F(1, 112) = 1.92$, $p = 0.16$, indicating that participants' improvement in performance on each of the grammaticality type was not influenced by the type of experiment (i.e.., L1 participants).

The descriptive statistics in Table 41 display the pattern of data relating to modality interventions indicating a greater mean improvement from the pre- to posttests in the auditory intervention modality ($M = 2.72$, $SE = 0.75$) compared to the visual intervention modality ($M = 1.89$, $SE = 0.61$). This pattern was present in some individual conditions (i.e., both items of grammaticality, and only items of TQ structures). However, analysis of participants' improvement in performance on NA and CON items reflected the opposite case. The ANOVA analysis, however, did not reveal a significant interaction between the testing session and modality, $F(1, 35) = 0.72$, $p = 0.40$, $\eta p^2 = 0.02$, demonstrating that the

improvement over data from auditory intervention was not significantly different from the visual modality interventions, as supported by the large overlap between the error bars of the two modality interventions in Figure 22 below. As a differential improvement effect was not observed between the two modality interventions on the EIT, there was no need for deeper analysis of relevant effects (e.g., to test whether any difference between the two interventions was influenced by type of syntactic structure or grammaticality).
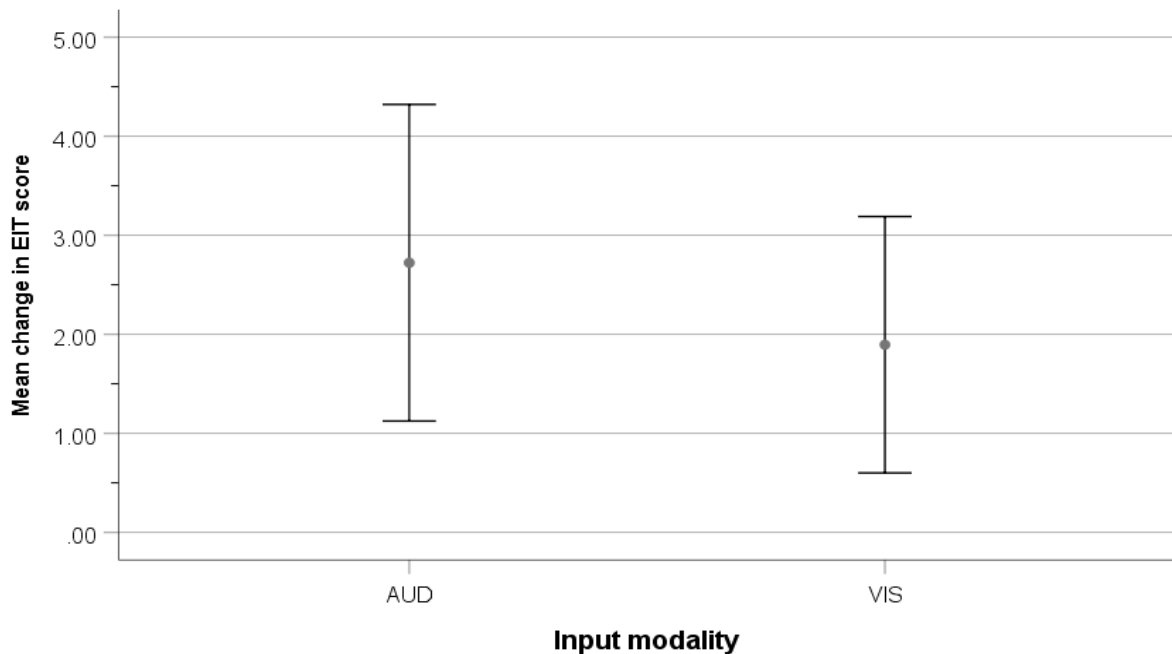
Figure 22. The difference in mean GJT improvement between auditory and visual modalities. Error bars indicate 95% confidence intervals.

Goal 2. The goal is to investigate whether auditory intervention participants' improvement in performance on each of the different syntactic constructions was influenced by grammaticality.

The results first revealed that there was a significant main effect of testing session, $F(1, 17) = 12.91$, $p = 0.002$, with a large effect size, $\eta p^2 = 0.43$, demonstrating an a positive

posttests ($M = 7.88$, $SE = 0.74$). Furthermore, the ANOVA analysis revealed that there was a significant interaction between the syntactic structure and grammaticality, $F(2, 34) = 8.38$, $p = 0.001$, with a large effect size, $\eta p^2 = 0.33$, demonstrating that the difference in the mean EIT scores across the three types of structure was influenced by the type of grammaticality. The analysis further revealed neither a significant interaction between the testing session and syntactic structure $F(2,34) = 0.09$, $p = 0.90$, $\eta p^2 = 0.00$, nor the testing session and grammaticality, $F(1, 17) = 1.54$, $p = 0.23$, $\eta p^2 = 0.08$. Moreover, and more importantly, the analysis revealed no significant interaction between the testing session, grammaticality and structure, $F(2, 34) = 0.57$, $p = 0.56$, $\eta p^2 = 0.03$, indicating that participants' improvement in performance on each of the different structures was not influenced by grammaticality.

Goal 3. The goal is to investigate whether visual intervention participants' improvement in performance on each of the different syntactic constructions was influenced by grammaticality.

The results first indicated there was a significant main effect of the testing session, $F(1, 18) = 9.46$, $p = 0.007$, with a large effect size, $\eta p^2 = 0.34$, reflecting a positive increase in the mean EIT score in the visual group from the pre- ($M = 5.10$, $SE = 0.68$) to posttests ($M = 7.00$, $SE = 1.00$). Moreover, the ANOVA analysis showed that there was a significant interaction between the syntactic structure and grammaticality, $F(2, 36) = 6.19$, $p = 0.005$, with a large effect size, $\eta p^2 = 0.25$, indicating the difference in the mean EIT scores across the three types of structure was influenced by the type of grammaticality. However, the analysis further revealed a significant interaction between the testing session and syntactic structure $F(2,36) = 4.15$, $p = 0.02$, with a large effect size, $\eta p^2 = 0.18$. Further analysis from the same mixed-ANOVA revealed that the first contrast comparing improvement in performance on the CON structure ($M = 1.15$, $SE = 0.36$) to the TQ structure ($M = - 0.21$, SE

= 0.40, the TQ was used as a baseline category), indicated a significant difference $F(1, 18) =$ 6.78, $p = 0.01$, with a large effect size, $\eta p^2 = 0.27$, demonstrating that participants' level of improvement based on learning was greater in the CON structure than in the TQ structure. The second contrast comparing improvement in performance on the NA structure ($M = 0.89$, $SE = 0.29$) to the TQ structure demonstrated that both structures did not elicit a significant difference in their patterns of improvement $F(1, 18) = 0.26$, $p = 0.61$, $\eta p^2 = 0.01$. These interactions are visualised in the interaction plot shown in Figure 23.
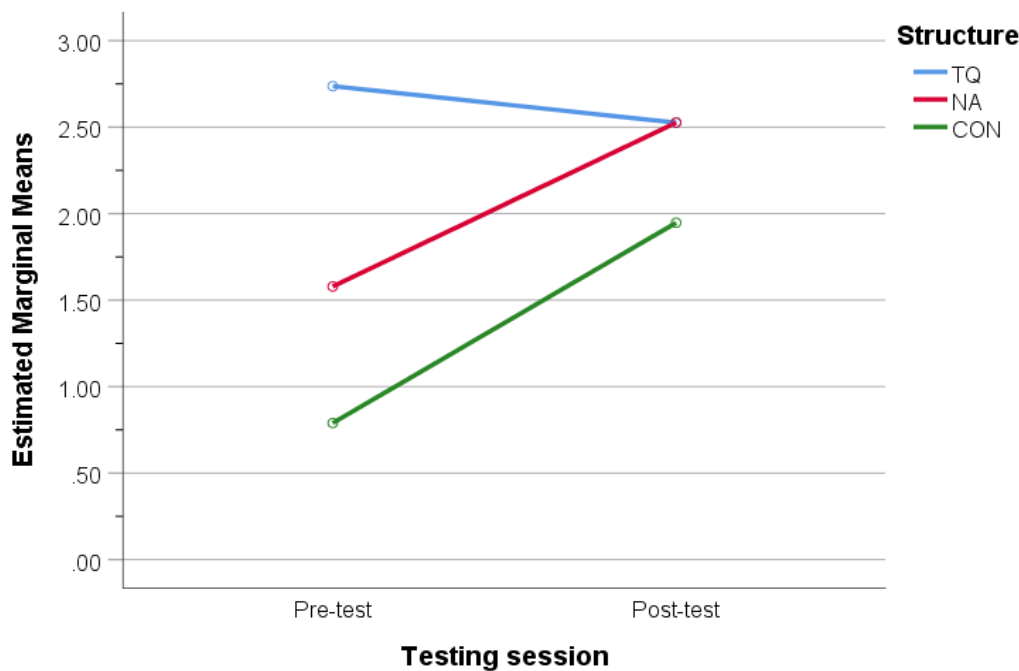

Figure 23. Performance for each structure type in the pre- and posttest by the visual group.

However, the analysis revealed no significant interaction between the testing session and grammaticality, but with slightly above chance, $F(1, 18) = 4.38$, $p = 0.051$, $\eta p^2 = 0.19$. More importantly, the analysis reported a significant interaction between the testing session, grammaticality and structure, $F(2, 36) = 3.46$, $p = 0.042$, with a large effect size, $\eta p^2 = 0.16$, indicating that participants' improvement in performance on each of the three distinct structures was influenced by grammaticality. Deeper analysis from the same ANOVA

revealed that the first contrast comparing improvement in performance on the NA structure to TQ structure (the TQ was used as a baseline category) differed significantly between grammatical and ungrammatical items, $F(1, 18) = 4.67$, $p = 0.044$, with a large effect size, $\eta p^2 = 0.20$. Similarly, the second contrast comparing improvement in performance on the CON structure to TQ structure differed significantly between grammatical and ungrammatical items, $F(1, 18) = 8.43$, $p = 0.009$, with a large effect size, $\eta p^2 = 0.31$. These interactions are visualised in Figures 24 and 25 below. The slopes of the lines representing the improvement in the TQ and NA structures for grammatical items were different to a similar extent for the same structures in ungrammatical items. Furthermore, the slopes of the lines representing the improvement in the TQ and CON structures for grammatical items were different to a similar extent for the same structures in ungrammatical items. It appears that the slope representing the TQ in the grammatical items was the casual factor for the distinct effects (i.e., the slope was steeper than the one in performance on ungrammatical items).
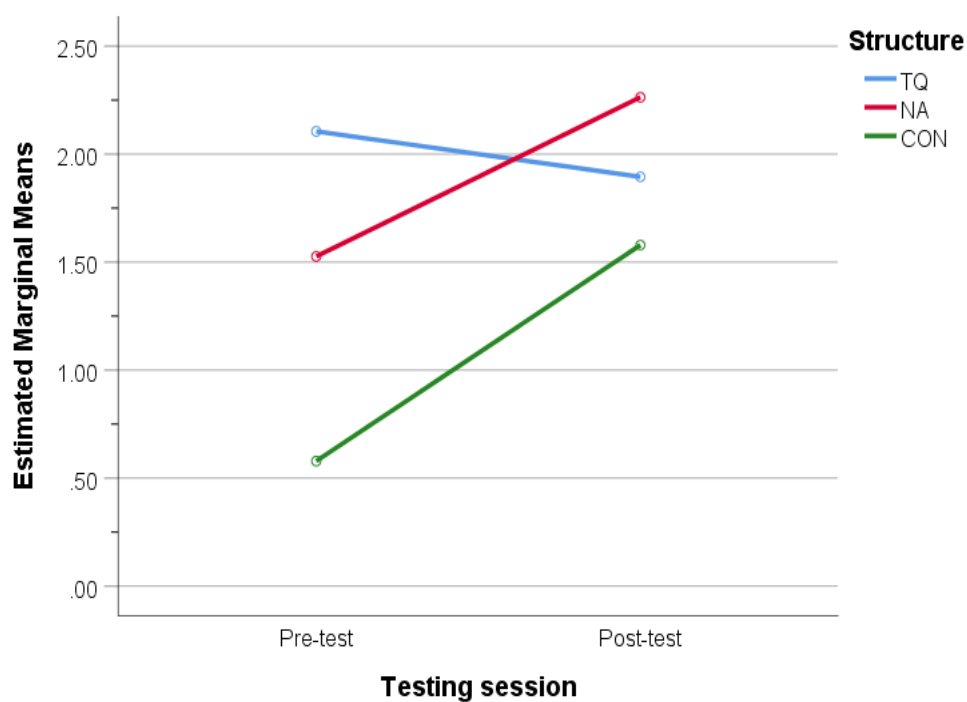
Figure 24. Performance for each structure type on grammatical items in the pre- and posttest by the visual group.
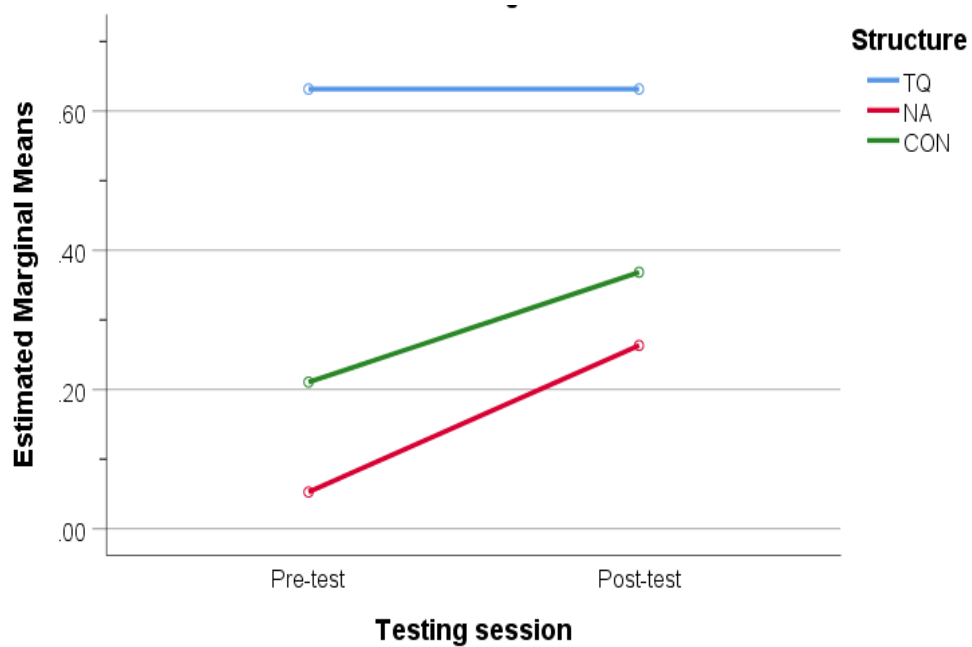

Figure 25. Performance for each structure type on ungrammatical items in the pre- and posttest by the visual group.

To summarise the results of analysis relating to the EIT, both auditory and visual exposure to L2 English syntax were found to result in successful acquisition of implicit grammatical knowledge. However, and more importantly, the results indicated no differentiable learning effects between visual and auditory interventions. For performance on syntactic structure across the two groups, no differing amounts of learning occurred for different syntactic structures, where similar results were present in the auditory group. In the visual group, however, the mean improvement in the TQ was significantly lower than that observed in the CON structures. The results further showed that the improvement in performance on grammatical items were greater than ungrammatical items. However, no significant difference in improvement was indicated between performance on grammatical and ungrammatical items within each modality group. Finally, participants' improvement in performance on each of the different structures was influenced by grammaticality in the visual group, but not in the auditory group.

Table 47

Summary of Mixed-ANOVAs for GJT and EIT Data Across Both Modality Groups

Variables	GJT			EIT		
	df	*F*	*P*	df	*F*	*P*
Main effects						
Testing session	1, 35	27.13	< 0.001	1, 35	22.58	< 0.001
Syntactic structure	2, 70	6.42	0.003	2, 70	27.93	< 0.001
Grammaticality	1, 35	186.38	< 0.001	1, 35	110.46	< 0.001
Interaction effects						
Testing × structure	2, 70	1.79	0.19	2, 70	2.27	0.11
Testing × grammaticality	1, 35	4.36	0.044	1, 35	5.08	0.03
Testing × modality	2, 35	1.80	1.88	1, 35	0.72	0.40

Table 48

Summary of Mixed-ANOVAs for GJT and EIT Data for Auditory Group

Variables	GJT			EIT		
	df	F	P	df	F	P
Main effects						
Testing session	1, 17	18.27	<0.001	1, 17	12.91	0.002
Interaction effects						
Testing × structure	2, 34	0.91	0.51	2, 34	0.09	0.90
Testing × grammaticality	1, 17	0.98	0.33	1, 17	1.54	0.23
Testing × structure × grammaticality	2, 34	1.00	0.37	2, 34	0.56	0.03

Table 49

Summary of Mixed-ANOVAs for GJT and EIT Data for Visual group

Variables	GJT			EIT		
	df	F	P	df	F	P
Main effects						
Testing session	1, 18	8.88	< 0.008	1, 18	9.46	0.007
Interaction effects						
Testing × structure	2, 36	0.77	0.46	2, 36	4.15	0.02
Testing × grammaticality	1, 18	3.94	0.06	1, 18	4.38	0.051
Testing × structure × grammaticality	2, 36	0.17	0.84	2, 36	3.46	0.42

5.6   Working memory analysis

The average score across the two experimental conditions on the OSPAN (a measure of WM capacity) was 54.35 ($SD = 12.16$, $SE = 2.00$) out of a possible total of 75. Participants in the auditory group scored an average of 53.00 ($SD = 14.81$, $SE = 3.49$), while those in the visual group scored an average of 55.63 ($SD = 9.22$, $SE = 2.11$). In the digit span (a measure

of phonological short-term memory), the mean score across the two experimental conditions was 5.84 (SD = 1.23, SE = 0.20) out of a possible total of 9. The mean score among the auditory group was 6.17 (SD = 1.42, SE = 0.33), while the mean score among the visual group was 5.53 (SD = 0.96, SE = .22). A Mann-Whitney U Test revealed that there was no significant difference in OSPAN scores between the auditory group (Md = 53) and the visual group (Md = 55.63), U = 166.00, z = – 0.15, p = 0.87. Similarly, there was no significant difference in scores on the digit span measure between participants in the auditory group (Md = 6) and those in the visual group (Md = 6), U = 138.00, z = –1.05, p = 0.29. Spearman's rank-order correlation coefficient was calculated over all participants and revealed no significant correlation between OSPAN and digit span performance, r = 0.18, p = 0.27. There was also no correlation when this test was carried out specifically among participants in the auditory group, r = 0.25, p = 0.31. However, there was no correlation between OSPAN and digit span scores among participants in the visual group, r = 0.13, p = 0.58.

Regression analysis by GJT across the groups

A simple linear regression was performed to investigate whether WM capacity (as measured by OSPAN score) was a significant predictor of overall improvement in performance on the timed GJT across the two experimental groups (auditory and visual). Both WM and the change scores from pre-to posttest were measured in points. The results indicated that WM was not a significant predictor of improvement in the timed GJT performance (p = 0.92). The model explained – 2.9 % of the variance and was not able to predict the improvement in GJT scores to a significant extent, $F(1, 35)$ = 0.01, p = 0.92.

To explore whether WM predicts posttest performance in the GJT when pretest and IELTS scores are controlled, A multiple linear regression was conducted where pretest and WM scores were measured in points, and IELTS scores were coded into three dummy variables: 6.5, 6.0, and 5.5 where the last band score was treated as a reference point. The

results indicated that only pretest was a significant predictor of posttest performance ($p = 0.1$), while WM, IELTS score: 6.0, and IELTS score: 6.5 were not. The model explained 14.6 % of the variance but was unable to predict posttest scores to a significant extent, $F(4, 35) = 2.49$, $p = 0.063$.

Regression analysis by GJT for auditory group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of improvement in performance on all GJT items for auditory group. The results indicated that WM was not a significant predictor of improvement in the timed GJT performance ($p = 0.85$). The model explained 6.4 % of the variance and was unable to predict the improvement in GJT scores to a significant extent, $F(1, 16) = 0.03$, $p = 0.85$.

To explore whether WM predict posttest performance in the GJT when pretest and IELTS scores are controlled, a multiple linear regression was conducted. The results indicated that none of the independent variables: pretest, WM, IELTS score: 6.0, and IELTS score: 6.5, was a significant predictor of posttest performance. The model explained 7.6 % of the variance and was unable to predict posttest scores to a significant extent, $F(4, 16) = 1.32$, $p = 0.31$.

As there was no significant relationship between WM and accuracy on GJT for the auditory group, there was no need for deeper analysis for the relationships between WM and grammaticality items, and items of syntactic structures.

Regression analysis by GJT for visual group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of improvement in performance on all GJT items for visual group. The results indicated that WM was not a significant predictor of improvement in the timed GJT performance ($p = 0.40$). The model explained -1.6 % of the variance and was unable to predict the improvement in GJT scores to a significant extent, $F(1, 18) = 0.72$, $p = 0.40$.

To explore whether WM predict posttest performance in the GJT when pretest and IELTS scores are controlled, a multiple linear regression was conducted. The results indicated that none of the independent variables: pretest, WM, IELTS score: 6.0, and IELTS score: 6.5, was a significant predictor of posttest performance. The model explained 22.2 % of the variance and was unable to predict posttest scores to a significant extent, $F(4, 18) = 2.28$, $p = 0.11$.

As there was no significant relationship between WM and accuracy on GJT for the visual group, there was no need for deeper analysis for the relationships between WM and grammaticality items, and items of syntactic structures.

Regression analysis by EIT across groups

A simple linear regression was further performed to investigate whether WM capacity was a significant predictor of improvement in performance on all items in the EIT across the two experimental groups (auditory and visual). The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.27$). The model explained 0.6 % of the variance and was not able to predict the improvement in EIT scores to a significant extent, $F(1, 35) = 1.22$, $p = 0.27$.

To explore whether WM predict posttest performance in the EIT when pretest, digit span (DS, as a measure of phonological short-term memory', and IELTS scores are controlled, a multiple linear regression was conducted. The results indicated that only pretest was a significant predictor of posttest performance ($p < 0.001$). The model explained 38 % of the variance and was able to predict posttest scores to a significant extent, $F(5, 35) = 5.28$, $p = 0.001$. Predicted posttest performance was equal to -0.09 + 0.63 (pretest score) - 0.14 (WM score) + 0.04 (DS score) + 0.41 (IELTS score: 6.0) - 0.06 (IELTS score: 6.5), where pretest, WM, and DS scores were measured in points, and IELTS scores were coded into three dummy variables: 6.5, 6.0, and 5.5 where the last band score was treated as a reference point.

As participants' pretest performance on the EIT increased by one point, their posttest performance increased by 0.63 points. As WM capacity increased by one point, the model's predicted EIT score in the posttest decreased by 0.14 points. As DS increased by one point, the model's predicted EIT score in the posttest increased by 0.04 points. A participant with an IELTS score of 6.0 was predicted to earn 0.41 more points on the EIT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to lose 0.06 more points on the EIT than a participant with an IELTS score of 5.5.

Table 50

Summary of Predictors of PostTest Performance in EIT Across Groups

	Model 1	
Variable	**Est.**	**SE**
Constant	- 0.09	(0.18)
Pretest	0.63***	(0.15)
WM	- 0.14	(0.14)
DS	0.04	(0.13)
IELTS (ref. 5.5)		
Score 6.0	0.41	(0.33)
Score 6.5	- 0.06	(0.36)
R^2	0.38	

Note. $n = 38$; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; *** $p < 0.001$.

Regression analysis by EIT for auditory group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of overall improvement in performance on the EIT for auditory group. The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.16$). The model explained 0.7 % of the variance and was unable to predict the improvement in EIT scores to a significant extent, $F(1, 16) = 1.11$, $p = 0.30$.

To explore whether WM predict posttest performance in the EIT when pretest, DS, IELTS scores are controlled, a multiple linear regression was conducted. The results indicated that none of these independent variables was a significant predictor of posttest performance. The model explained 17.9 % of the variance and was unable to predict posttest scores to a significant extent, $F(5, 16) = 1.69$, $p = 0.21$.

Regression analysis by EIT for visual group

A simple linear regression was performed to investigate whether WM capacity was a significant predictor of overall improvement in performance on the EIT for visual group. The results indicated that WM was not a significant predictor of improvement in the EIT performance ($p = 0.96$). The model explained -5.9 % of the variance and was unable to predict the improvement in EIT scores to a significant extent, $F(1, 18) = 0.002$, $p = 0.96$.

To explore whether WM predict posttest performance in the EIT when pretest, DS, IELTS scores are controlled, a multiple linear regression was carried out. The results of a multiple linear regression indicated that only pretest was a significant predictor of posttest performance ($p < 0.001$). The model explained 60.5 % of the variance and was able to predict posttest scores to a significant extent, $F(5, 18) = 6.52$, $p = 0.003$. Predicted posttest performance was equal to 0.12 + 1.03 (pretest score) - 0.06 (WM score) - 0.01 (DS score) - 1.07 (IELTS score: 6.0) - 0.47 (IELTS score: 6.5). As participants' pretest performance on the EIT increased by one point, their posttest performance increased by 1.03 points. As WM capacity increased by one point, the model's predicted EIT score in the posttest decreased by 0.06 points. As DS increased by one point, the model's predicted EIT score in the posttest decreased by 0.01 points. A participant with an IELTS score of 6.0 was predicted to lose 1.07 more points on the EIT than a participant with an IELTS score of 5.5. A participant with an IELTS score of 6.5 was predicted to lose 0.47 more points on the EIT than a participant with an IELTS score of 5.5.

Table 51

Summary of Predictors of Posttest Performance in EIT for Visual Group

Variable	Model 1	
	Est.	SE
Constant	0.12	(0.21)
Pretest	1.03***	(0.19)
WM	- 0.06	(0.23)
DS	- 0.01	(0.25)
IELTS (ref. 5.5)		
Score 6.0	- 1.07	(0.65)
Score 6.5	- 0.47	(0.44)
R^2	0.60	

Note. n = 38; WM = working memory; IELTS = International English Language Testing System; Est. = estimate. SE = standard error. R^2 = adjusted R-squared; *** $p < 0.001$.

5.7 Discussion

Experiment 2 was conducted to examine to what extent the main effects and relationships observed in Experiment 1 were consistent with the results of Experiment 2 when carried out with participants who had a different native language. Similar to Experiment 1, Experiment 2 investigated the effects of manipulating the input modality with L1 Arabic speakers to determine the extent to which exposure to L2 syntax in each modality (auditory or visual) can stimulate automatic processing in the L2 system, resulting in the successful acquisition of implicit grammatical knowledge (Research Question 1). The experiment further explored whether WM capacity predicts how this automatic processing is acquired differently with exposure to the two different modalities (auditory vs. visual, Research Question 2).

5.7.1 Modality

The results of data across the modality groups demonstrated a significant mean improvement from the pre- to posttest, but with a slight increase in the mean accuracy.

However, because this slight increase was determined by both measures (i.e., EIT and timed GJT), one might be more confident that the slight increase is not due to fewer 'not sure' responses (in the case of timed GJT), a task practice effect, or a change in task strategy after training.

The results provided further evidence that exposure in either modality intervention (auditory or visual) can boost the automaticity of syntactic processing, as determined by both the timed GJT and the EIT. However, similar to the results elicited from Experiment 1, the increase in the level of automatization from pre-to posttests did not differ significantly between the two modality interventions as determined by both measures.

In addition, no difference was found in the pretest scores between the two modality groups, suggesting that the lack of a differential learning effect was not affected by any pre-existing differences between the two modality groups.

The results of Experiment 2 are inconsistent with the general theoretical claims in terms of the difference in the encoding and processing modality. Based on the literature review, the visual stimuli must undergo articulatory rehearsal to be encoded in a phonological format and then transferred into the phonological store. In contrast, the auditory material gains direct and automatic access to the phonological store and does not need to undergo such re-encoding. Therefore, it was assumed that the rehearsal process involved in processing visual stimuli recovers the memory trace and, as a result, the decay of the trace is repressed (Baddeley, 1990). Furthermore, on a superficial level, the nature of each modality has its inherent idiosyncrasies, which were assumed to be ascribed to the difference in processing. In the auditory case, the variation in prosodic signals, rapid speed of input, and high degree of transience impose heavy demands on processing resources, which most likely has detrimental effects on successful auditory comprehension. This is unlike the processing of the visual modality, which allows for greater permanency (i.e., recursive processing) and control over

the input rate in addition to the stability inherent in its orthographic linguistic representation, resulting in reducing the incidence of false learning and recognition and deeper processing of linguistic information.

Empirically, the results are not in line with Experiment 1, which found an observable trend towards a visual modality advantage in grammatical knowledge automatization. Similarly, the results of Experiment 2 are inconsistent with those of Johnson (1992) and Murphy (1997), who found lower accuracy in metalinguistic tasks presented in the auditory modality (i.e., untimed GJT); of Park (2004), who found poorer comprehension in the auditory task; or of Wong (2001) and VanPatten (1990), who found poorer comprehension in the auditory modality when attention is focused on meaning and form simultaneously.

However, several previous studies (e.g., Morgan-Short et al., 2018; Sagarra & Abbuhl, 2013; Sydorenko, 2010) failed to find any differential effect in favour of visual material, which is in line with the results of Experiment 2, thus contradicting the general claims about the advantageous bias to the visual modality in terms of language development. Sagarra and Abbuhl (2013), for instance, employed language tests which were biased in favour of the use of explicit knowledge; they found that auditory and visual feedback did not differ significantly in L2 learning as determined by written tests. However, they found counter-evidence in aural tests (i.e., advantageous bias to the auditory feedback). In this case, one cannot argue that the similar findings observed in both Experiment 2 and Sagarra and Abbuhl (on written tests, 2013) are a result of the lack of difference in the encoding and processing modality between the auditory and visual modalities.

Because Experiments 1 and 2 involved identical training and testing procedures, and the former revealed a clear tendency towards a visual modality advantage while the latter did not, this suggests that other factors might play a role in generating different results between Experiments 1 and 2, such as the small number of participants in Experiment 2.The small

sample size can affect ANOVAs and correlations alike (i.e., lack of statistical power). More specifically, 'Type 1' errors (i.e., assuming there is an effect when there isn't) or 'Type 2' errors (i.e., assuming there isn't an effect when there is) may occur.

Another factor may account for the difference in results between the two experiments relates to prior available automated knowledge. The results demonstrated that participants in Experiment 1 performed significantly better in the pretest than did participants in Experiment 2, as determined by the GJT, and it was close to being statistically significant in the EIT scores ($p = 0.07$). Due to the assumption that this measure assesses the automatic processing of L2 grammatical knowledge, low proficiency in implicit linguistic knowledge (i.e., the application of attentional processes is maximised) might be the causative factor beyond the lack of difference in the development of automated knowledge under the two modality interventions in Experiment 2. This could be evidenced by improvement across the whole sample in performance on grammatical and ungrammatical items, where the former elicited a significantly higher average score than the latter, as established by the GJT and the EIT. Thus, it is safe to hypothesise that there is an intersection between the learners' readiness to increase the automated knowledge of a given structure and the modality in which this structure is presented. This hypothesis, while preliminary in the implicit knowledge acquisition, can be supported by (Mackey & Philp, 1998), who found that only advanced learners (i.e., at the required stage of L2 acquisition) were able to gain an advantage from recast in the use of English question forms (see Section 7.2 for more information about developmental readiness hypothesis).

5.7.2   Syntactic structure

Across the whole sample, the amount of improvement did not vary across the three structures, that is, TQ, NA, and CON, as shown in both measures, namely, the timed GJT and the EIT. Similar results were present in the data from the auditory group. However, the

results revealed that the variation in the amount of improvement across the structures was observed in the visual modality group as determined by the EIT.

To examine how this variation occurred across the structures in the EIT, analysis indicated that, of the three structures used in the stimuli, the TQ items elicited the highest mean score in the pretest performance. This result was in evidence in the overall performance on the pretest in the timed GJT. This suggests that participants had fairly good knowledge of the TQ structure prior to their exposure to the intervention in comparison to the NA and CON structures. However, the TQs elicited a small average decrease at the posttest ($M = -0.21$). The possible explanation for the lack of improvement refers to the increased average score at the pretest and the two TQ-inherent factors influencing conceptual complexity (i.e., complex function and difficult processing, see more discussion in Experiment 1). A similar performance was also observed in Experiment 1 where the TQs elicited the lowest mean improvement score in the EIT across the groups and in each modality group.

Unlike the TQs, analysis of the data pertaining to the NA and CON structures reflected a high pattern of improvement. However, the former did not reflect a significant distinct improvement from that observed in the TQs, but the latter did. The results are in line with Experiment 1, where both experiments revealed that the CON structure reflected a district mean improvement from that observed in the TQ structure. However, the difference between the two experiments relates to the NA structure where such a structure in Experiment 2 reflected a distinct improvement from that observed in the TQ across the whole sample and in each modality group.

An explanation that might account for the results of Experiment 2 relates to the poorest performance the CON in the pretest, which could help participants make a greater improvement than in the NA structure. This strong improvement was observed despite three factors contributing to conceptual complexity being in operation in the rule system governing

the CON structure, specifically, its perceptual non-salience, irregularity, and functional complexity. This might be explained by the nature of the embedded inversion, which is achieved by replacing the subject with the auxiliary and omitting the adverbial subordinator *if*, as in *Had you worked hard in your bachelor's studies, you would have gained higher grades*. Participants appeared to have already had basic knowledge of the conceptual syntax governing the normal word order of *if*-conditional sentences, as in *If you had worked hard in your bachelor's studies, you would have gained higher grades*. However, subject-operator inversion as a result of omitting *if* was probably the cause of subjects' difficulty in comprehending the rule system. Thus, following the intervention stage, in which rule instruction for the CON structure was provided, participants became aware of the subject-operator inversion that represented a clue in uncovering the rule system.

5.7.3 Grammaticality

The increase in performance from the pre- to posttests across the two modality groups was affected by the type of grammaticality as indicated by the GJT and the EIT. This means that the overall improvement in performance on grammatical items was greater than improvement in performance on ungrammatical items. Similar results were observed in the visual group only when improvement in performance on the different structures were analysed based on grammaticality. The results revealed the comparison of improvement in performance on the NA and CON structures to the TQ structure differed significantly between grammatical and ungrammatical items (i.e., improvement in the TQ structure was noticeably lower for grammatical items than that observed for ungrammatical items, see Figure 24 and Figure 25). However, the differential effects of grammaticality were not evident when the data from auditory and visual groups were analysed separately (the difference in the visual group was slightly above chance in the EIT).

The results generally suggest that participants' L2 syntax in Experiment 2 was not automatized to such an extent that ungrammatical items were correctly rejected in the case of the GJT or generated oral fluent correction of syntactically deviating elements in the case of the EIT. This is unlike Experiment 1, which found no difference in performance between grammatical and ungrammatical items. These results support the interpretation mentioned above that a difference in participants' prior automated knowledge might have played a distinct role in terms of the type of grammaticality.

The results support the claim that responses to grammatical and ungrammatical sentences draw on distinct sources of knowledge. Ellis (1991, p. 178) observed that "sentences that learners judged to be ungrammatical or that they were not sure about often invoked attempts to make use of declarative knowledge". The results are compatible with those of Ellis (2005), Gutiérrez (2013), and J.-e. Kim and Nam (2017), who found a distinction between learners' responses to grammatical and ungrammatical items on the timed and untimed GJT. Ellis (2005) found that participants' scores on ungrammatical items loaded more strongly on an explicit knowledge factor than their scores on grammatical items, as shown in the untimed GJT. Gutiérrez's (2013) research indicated a clear distinction between learners' responses to grammatical and ungrammatical strings in both the timed and untimed GJTs, suggesting that implicit knowledge is accessed when judging grammatical strings, while explicit knowledge is used when judging ungrammatical ones. However, J.-e. Kim and Nam (2017), in a study that presented the timed GJT in the auditory modality, discovered that responses to ungrammatical sentences loaded more strongly on an implicit knowledge factor than grammatical sentences.

The results are also consistent with what has been found in Reinders and Ellis's (2009) study, in which overall learning effects on the acquisition of implicit knowledge of the NA structure were observed only when grammatical and ungrammatical sentences were

tested separately, as determined by the timed GJT. That is, the overall improvement in performance of grammatical items was significantly greater than that observed of ungrammatical items. The results of Experiment 2 and Reinders and Ellis's (2009) study suggest the knowledge of L2 syntax was not automated to a sufficient extent, and consequently, more cognitive demands were exerted to process an ungrammatical item. This means that when explicit knowledge pertaining to the syntactically deviating element needs to be retrieved, the time pressure inherent in the timed GJT and the EIT suppresses explicit monitoring, resulting in detrimental effects on performance.

Ellis and Roever (2018, p. 11) concluded that "clearly, results for grammatical versus ungrammatical sentences have not been consistent across studies although the weight of the evidence supports Ellis' finding, namely that ungrammatical sentences are the more likely to elicit explicit knowledge". However, one could argue that while responses to grammatical sentences measure subjects' implicit knowledge, L2 learners' ability to reject ungrammatical items under time constraints in the GJT or their ability to generate oral fluent correction of syntactically deviating elements in the case of the EIT can be interpreted as a purer measure of implicit knowledge, as supported by J.-e. Kim & Nam (2017).

To summarise the discussions relating to the first research question, Experiment 2 provides further evidence that training in either modality intervention (auditory or visual) is able to foster the automaticity of syntactic processing, as measured by both the timed GJT and the EIT. However, the results demonstrated no differentiable learning effects between the two modality interventions. This might be due to the restricted implicit knowledge of L2 syntax that participants had in the beginning as determined by the pretest scores; however, more evidence is needed for a conclusion to be drawn. With regard to syntactic structure, the amount of language improvement varied across the three structures as determined by the EIT in the visual-only group where the CON structure reflected a greater improvement than that

observed in the TQ structure. For grammaticality, across the whole sample, performance on grammatical items improved more than that observed in ungrammatical items in both the GJT and the EIT. This probably suggests that the knowledge of L2 syntax was not automated to such an extent that participants correctly responded to ungrammatical items.

5.7.4   Working memory

The results revealed that the effects of neither auditory nor visual input on the development of L2 automatic processing, as determined by both EIT and timed GJT, were mediated by WM capacity. This result, specifically the poor relationship between WM and EIT in both modality groups, can be further supported by the lack of relationship between phonological short-term memory (as measured by a digit span task) and the EIT performance.

The results are not surprising and are in line with the expectations of the findings of the first research question. Given that a similar pattern of improvement from pre- to posttests was found between the two modality interventions, it is sensible that WM capacity was not found to correlate with language development accrued from any type of modality exposure. This interpretation can be supported by the results of Experiment 1; as the results reflected a clear trend towards a visual modality advantage in language learning, WM capacity did mediate the effects of auditory modality on language development.

Moreover, the results of Experiment 2 are inconsistent with the general theoretical claims regarding the difference in the computational demands imposed by different input modalities on WM capacity. Auditory linguistic input is assumed to exercise greater processing demands on WM capacity than visual linguistic input, and this is largely due to the distinctive idiosyncrasies inherent in the nature of auditory and visual processing (see Experiment 1 for more discussion).

Compared with the previous empirical studies, the obtained results broadly contradict the major trends of previous studies investigating the relationship between WM and L2 visual

performance (e.g., Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Leeser, 2007; Walter, 2004). Similarly, the results are incompatible with those of Kormos and Sáfár (2008), who found that both L2 auditory and visual comprehension were predicted by WM capacity. However, the results are partly in line with previous empirical studies on the relationship between WM and L2 auditory comprehension (e.g., Andringa et al., 2012; Brunfaut & Révész, 2015; Vandergrifta & Baker, 2015). Brunfaut and Révész (2015) did not find significant correlations between WM capacity and the listening test designed to measure L2 listeners' global comprehension. Moreover, Andringa et al. (2012) and Vandergrifta and Baker (2015) found that WM capacity is a weak predictor of auditory comprehension when other factors are controlled.

More importantly, the results did not parallel evidence from Sagarra and Abbuhl (2013), who found WM capacity is correlated with the amount of learning accrued from auditory feedback but not from visual feedback, based on language tests (written and oral). The explanation, which most probably accounts for the lack of any relationship in WM capacity–auditory modality in Experiment 2 is due to the fact that the level of participants' knowledge of L2 syntax was not automated to the extent that L2 learners with high and low WM capacity can be distinguished due to the efficiency of their automatic processing as reflected in the timed GJT and the EIT. However, more evidence is needed for a conclusion to be drawn.

To summarise the discussions relating to the second research question, the results revealed that WM capacity did not interact with input-based language learning for either modality, as determined by the timed GJT and the EIT. The results did not support the claim that the processing demands exerted by auditory linguistic input on WM capacity are greater than those exerted by visual input, leading the former to distinguish between individuals with high and low WM capacity due to the efficiency of their auditory input processing. Similarly,

the results are not in line with either Experiment 1 or Sagarra and Abbuhl (2013), both of which found that a relationship between WM capacity and language development occurred from the auditory input. More importantly, the obtained data pertaining to the first and second research questions provided relatively cohesive results; that is, as no modality intervention was indicated as being more effective than another at fostering L2 automatic processing, no modality intervention, as determined by the timed GJT and the EIT, was found to correlate with WM capacity.

Chapter 6: General discussion

The work described in this thesis was conducted to investigate two factors affecting the process of acquiring implicit knowledge of L2 syntax (reflected in automatic processing). The first is concerned with the modality of stimulus presentation (auditory vs. visual) during training, and the second is related to the role of WM capacity in the acquisition of this implicit knowledge. The following discussion will be developed by means of a summary of the study's motivation, a summary of results, and consideration of the outcomes of Experiments 1 and 2. Next, the focus of the chapter shifts to discuss the difference in performance between the timed GJT and the EIT, addressing the underlying variables operationalised by both measures. Finally, in the conclusion section the chapter will discuss the theoretical and experimental implications and address the limitations of the study.

6.1   Summary of motivation

The phenomenon by which input modality can affect language acquisition was of specific interest, based on previous research that has indicated that auditory and visual inputs are represented and processed in two separate streams, with each stream possessing its own properties (Penney, 1989). On a superficial level, visually presented language allows for repeated processing and permits the reader to control the rate of input, whereas auditory presentation does not allow this, largely as a result of the rapid nature of the speech stream and its transience (i.e., the fact that the input stream cannot be re-inspected). In auditory processing, a processor in the human brain must further deal with a variation in prosodic signals (i.e., phonological recognition of words, unit boundaries, stresses, and intonational patterns) as prerequisite clues for linguistic comprehension. This is unlike visual processing, where the processer parses the input in the absence of elements related to prosody; that is, representations of visual language are orthographically immutable. Representative evidence to support the modality differences mentioned above comes from psychological and L1

research on the effects of input modality on comprehension and performance processing (e.g., Dennis, 1977; Penney, 1989, Danks, 1980; Anderson, 1980; Shaffer, 1975). Their line of research identified advantages for the visual modality. Additional empirical evidence has appeared in L2 research investigating the distinctive effects of input modality on L2 comprehension and performance (e.g., Johnson, 1992; Murphy, 1997; Park, 2004; VanPatten, 1990; Wong, 2001).

The unique idiosyncrasies of each modality together with the results of previous L1 and L2 research mean that it is reasonable to hypothesise there will be a distinction between the two modalities in terms of how such input exerts an influence on the development of the automatic L2 linguistic system. However, research investigating the effects of input modality on L2 learning itself is very limited (e.g., Sagarra and Abbuhl, 2013; Sydorenko, 2010); the existing results have been mixed, and some tests have revealed counter-evidence (i.e., an advantageous bias towards auditory input).

Therefore, the first research question of the work reported on in this thesis pertained to the extent to which exposure to L2 syntax in each modality (auditory or visual) can stimulate automatic processing in the L2 system, resulting in the successful acquisition of implicit grammatical knowledge. Unlike the previous studies that investigated the effects on L2 language development of manipulating modality, the present investigation relied on an EIT and a timed GJT for assessing language development, as these are assumed to measure implicit knowledge (Bowles, 2011; Ellis, 2005, Erlam, 2006; Loewen, 2009; J.-e. Kim & Nam, 2017; Loewen, 2009; Suzuki & DeKeyser, 2015; Zhang, 2015). Both of these tasks involve a certain level of automaticity, and hence, such assessment was believed to elucidate whether the L2 linguistic system is susceptible to becoming automated via auditory or visual input-based language learning.

The EIT is a demanding task that involves both auditory input and oral output processing. Such a measure requires a participant to reproduce utterances that he or she has heard while processing them for semantic meaning and correcting any syntactic anomalies, under time pressure (please refer to the methodology chapter for more information on how participants' attention was attracted to the utterance's meaning). The underlying hypothesis is that spontaneous and fluent correction of deviant structures is a strong and sufficient indication that automatic language processing is in operation (i.e., implicit knowledge has been successfully acquired and can now be accessed, Erlam, 2009).

The GJT, in contrast, requires a participant to judge whether a visually presented sentence is grammatically well formed or deviant under time constraints. Although the nature of such a task requires a sentence to be processed for its linguistic form, strict time pressure is hypothesised to trigger access to implicit knowledge, if available (Loewen, 2009). Put differently, the likelihood of the participant being able to reprocess the stimulus and actively contemplate his or her response is greatly minimised, and the likelihood of exercising his or her intuitive linguistic judgement is increased, meaning that a superior performance indicates a high level of automaticity in the participant's acquired grammatical knowledge.

In light of the modality differences discussed above, the acquisition success of L2 learners' implicit knowledge can be affected not only by differences in modality presentation but also by WM capacity. WM capacity describes to what extent the amount of information can be stored and processed simultaneously over a given temporal interval. The predictive power of WM capacity on a wide range of phenomena in second language was assumed to arise from the greater demands placed on cognitive resources by WM operations carried out during L2 processing compared to during L1 processing due to a lack of high-level competence, including automaticity (see Wen, 2015 for a review). More specifically, due to the intricate processing required to decode and parse an L2 auditory speech stream (e.g., as a

result of the varied prosodic patterns and the transient and rapid nature of the input), the linguistic processing demands exerted by these factors should be expected to tax the WM resources heavily. In this respect, high WM capacity is most likely to provide a greater benefit to learning than low WM capacity based on auditorily presented language. This is distinct from the case in the visual stream, which should involve a lower level of linguistic demands on processing resources in WM, and consequently, WM is likely not to mediate the effects of visual input on the development of L2 automatic processing for low or high WM capacity.

Previous research has provided empirical evidence to indicate that WM capacity is favourable for L2 visual performance (e.g., Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Leeser, 2007; Jeon & Yamashita, 2014), while evidence on the WM capacity-L2 auditory relationship is inconclusive (e.g., Kormos and Safar 2008, Miyake and Friedman, 1998; Andringa et. Al, 2012; see Sakai, 2018 for review). This might be due to the dearth of research between the two variables. These previous studies failed to elicit sufficient evidence to support the general theoretical claim that computational demands exerted by auditory language comprehension on WM are greater than those exerted by visual comprehension. To the best of the author's knowledge, only one study ((Sagarra & Abbuhl, 2013) has addressed the relationship between WM capacity and the amount of learning accrued from the two different input modalities (auditory vs. visual) and found results that corroborated the general claims. That is, the effects of auditory, not visual input, on the development of L2 learning were mediated by WM capacity. Despite the successful work in Sagarra and Abbuhl's (2013) study in terms of WM capacity, the substantial challenge remains regarding whether WM capacity still interacts with the amount of learning accrued from auditory input when this amount of learning is gauged through the EIT and the timed GJT, designed as measures of primarily implicit grammatical knowledge (i.e., L2 acquisition).

Thus, the second research question of this thesis was whether WM capacity interacts with L2 automatic processing development and training in two input modalities (auditory and visual). The WM capacity was measured by an OSPAN task, which is considered a complex task tapping into multiple mechanisms and functions pertaining to the whole cognitive construct (i.e., maintenance and executive control). The OSPAN task is designed in such a way as to induce activation of both the storage component and the processing component of memory during the assessment. In this task, the participant is required to verify simple mathematical equations, following each of which a letter is displayed. The participant is then asked to recall the letters that followed each mathematical equation in the order presented. This task differs from the digit span task, a measure of phonological short-term memory, which involves auditory exposure to only a set of digits followed by immediate recall in order. The latter type of task, which requires only temporary storage of information, was employed in the present study only as a control measure for the EIT because the latter involves oral repetition over a given temporal interval.

6.2   Summary of results

Chapter 4 presented an experiment in which the first language of the participants was Chinese. The participants were randomly assigned to one of the two modality interventions (auditory and visual), and they attended two sessions. In the first session, they first completed the digit span task (a measure of short-term memory) and the OSPAN task (a measure of WM). Then they were pretested using the timed GJT and the EIT. In the second session, both experimental groups had been exposed to identical training (i.e., focus on form and explicit learning tasks) in the three structures (i.e., TQ, NA, and CON) where the only difference between the groups is the modality of stimuli presentation. Finally, a posttest similar to the pretest was immediately administered but with a different version. The results demonstrated no significantly differentiable effects between visual and auditory interventions, but showed

better learning was clearly observed in the visual intervention. As a result, the difference in

improvement between the two interventions was marginally significant in both the GJT and

the EIT measures ($p = 0.052$ and $p = 0.08$, respectively), where both of them reflected a small

effect size ($\eta p^2 = 0.05$ and $\eta p^2 = 0.03$, respectively). Similar results elicited from the two

measures (i.e., timed GJT and EIT) should be viewed as the degree of consistency between

the two. Neither syntactic structure nor grammaticality affected the possible difference

between the two modality interventions differently in either task. Moreover, no significant

difference between the two groups was indicated by pretest scores at the outset of the

experiment. These two results can be interpreted as indicating that the clear tendency towards

better learning via visual input was most likely due to the visual intervention itself and not

due to pre-existing differences between the two groups or differences in performance on

grammatical or syntactic items between the groups.

In relation to performance on syntactic structures and type of grammaticality, analysis

of accuracy in the timed GJT revealed that across the two modality groups, a similar

improvement was observed for all three types of syntactic structure and the two types of

grammaticality. A similar result was further found when the data from the auditory and visual

groups were analysed separately. However, analysis of accuracy in the EIT showed that the

overall amount of improvement varied for different syntactic structures, where the mean

improvement for the TQ was significantly lower than the NA and CON structures. This

pattern of results was further present in the visual group. In the auditory group, however, the

difference only lay between TQ and NA structures. For grammaticality, the results revealed

no significant difference in improvement between overall performance on grammatical and

ungrammatical items in the EIT. This lack of difference was further present within each

group. In terms of the interaction between syntactic structures and the type of grammaticality,

participants' mean improvement in performance on the three distinct structures was

influenced by grammaticality, as determined only by the EIT. In particular, improvement in the CON structure was noticeably lower for ungrammatical items than that observed for grammatical items whereas the opposite was observed for the TQ.

In relation to WM capacity, as measured by the OSPAN task, was found to correlate positively with overall improvement in performance following only auditory exposure in the GJT but not in the EIT. Deeper analysis of the GJT data demonstrated that this correlation appears to be driven by the correlation between OSPAN and accuracy in ungrammatical items and the correlation between OSPAN and accuracy in NA items. Similar results were obtained when IELTS scores were taken into account using multiple regressions.

Chapter 5 presented an experiment in which the first language of participants was Arabic. The main goal of Experiment 2 was to verify the findings in Experiment 1 but in a different language group. First, the results provided increasing evidence that training in either modality intervention (auditory or visual) can promote the automaticity of syntactic processing as measured by both the timed GJT and the EIT. However, the results revealed no significantly different learning effects between the visual and auditory interventions.

In relation to performance on syntactic structures and type of grammaticality, the analysis of accuracy in the timed GJT indicated that the overall language improvement did not vary across the syntactic structures. Similar results were further found within each modality group. However, a significant increase in performance on grammatical items was greater than in ungrammatical items across the two groups, but below chance ($p = 0.044$). This significance was not present within each modality group. Analysis of accuracy in the EIT, on the other hand, showed that the overall amount of improvement did not vary for different syntactic structures across the two modality groups. However, the variation was shown only within the visual group, where the improvement in performance on the CON structures was significantly greater than on the TQ. For grammaticality, the results revealed a

significant difference in improvement between performance on grammatical and ungrammatical items, where the average score for the former was higher than the latter. However, this significance was not shown within each modality group. In terms of the interaction between syntactic structures and the type of grammaticality, the results revealed that type of grammaticality was found to affect participants' mean improvement in performance on the three distinct structures, as determined only by the EIT. In particular, improvement in the TQ structure was noticeably lower for ungrammatical items than that observed for grammatical items.

For WM capacity, as measured by the OSPAN task, the analysis revealed that neither the effects of auditory nor of visual input were found to correlate with WM capacity on the development of L2 automatic processing.

6.3   Joint consideration of the outcomes of Experiments 1 and 2

According to both Experiments 1 and 2 as reported here, the results related to the first research question, the extent to which exposure to L2 syntax in each modality (auditory or visual) can stimulate automatic processing in the L2 system, reflected no significant difference in improvement between the auditory and visual modality interventions. However, Experiment 1 revealed an apparent tendency for the process of automating L2 grammatical knowledge to be fostered via the visual modality more successfully than by the auditory modality, as determined by both implicit knowledge measures, the GJT and the EIT. Although the analysis in Experiments 1 and 2 indicated no significant difference between the two modality groups on pretest scores at the outset of each experiment (i.e., the obtained results at posttest were not influenced by pre-existing differences between the two modality groups), the difference in obtained results between Experiments 1 and 2 could be accounted for by the difference in participants' available automated knowledge of their English syntax prior to exposure to treatment interventions. Put differently, analysis demonstrated that

participants in Experiment 1 performed significantly better in the pretest than participants in Experiment 2, as determined by the GJT. If the observed tendency towards visual modality advantage was taken into serious consideration, low proficiency in implicit grammatical knowledge (i.e., the application of attentional processes is maximised) might be the causative factor beyond the lack of difference in the development of automated linguistic knowledge between the two modality interventions in Experiment 2. This could be evidenced by improvement across the whole sample in performance on grammatical and ungrammatical items, where the former elicited a significantly higher average score than the latter, as established by the GJT and the EIT. This suggests that the knowledge of L2 syntax was not automated to such an extent that participants were able to correctly reject ungrammatical items in the case of the GJT or generate oral fluent correction of syntactically deviating elements in the case of the EIT. This is unlike Experiment 1, which did not find a significant difference in improvement between performance on grammatical and ungrammatical items across the whole sample and in each modality group, as determined by both measures: the EIT and the timed GJT). If this is the case, then the results of the present study provide important information regarding the interaction between prior proficiency in implicit linguistic knowledge and modality-based learning. That is, which level of implicit knowledge—low, intermediate, or upper-intermediate—is more sensitive to stimulating automaticity under auditory and visual modality-based language learning? (see Section 7.2 for more discussion)

When the results of the two experiments are juxtaposed, some consistency of conclusions can be drawn. Improvement in performance did not vary across the syntactic structures, as determined by the GJT in both experiments. However, the amount of improvement did vary across the structures in the EIT for both experiments. Experiment 1, conducted with native speakers of Chinese, revealed a significantly superior improvement in

performance of the NA and CON structures than the TQ structure, where the latter reflected a very small average increase in score between the pre- and posttests (the TQ was used as a baseline category in the ANOVA). Similar results were observed in the visual group, but in the auditory group, only performance of the NA structure improved more than the TQ structure. In Experiment 2, conducted with native speakers of Arabic, there was a difference in the amount of improvement across structures between the CON and the TQ structures, with a greater mean improvement observed in the CON than in the TQ.

The results determined by the EIT from both experiments provide further evidence that the poor increase in performance of the TQ structure from pre- to posttests is the principal causative factor in generating a statistically significant difference in improvement across the syntactic structures. Two possible factors can account for this low improvement: prior knowledge of the TQ structure and difficult processing. Pretest data revealed that the TQ structure elicited the highest mean score in both tasks (the timed GJT and the EIT) for both experiments, suggesting that participants had a relatively thorough knowledge of the TQ structure prior to their exposure to the training intervention. As a result, participants might have found it difficult to obtain a further increase in performance at the posttest. From a psycholinguistic perspective, processing this structure is most likely to demand cognitive effort due the requirement for procedures pertaining to the processing of a subordinate clause. More specifically, TQ processing requires that the appropriate auxiliary form is entirely governed by the verb phrase in the independent clause, which increases the competence complexity of the syntax governing the TQ rule.

Despite the processing idiosyncrasy inherent in the syntactic complexity of the TQ rule, this structure includes two factors that contribute to its conceptual simplicity: a sentence-final position and regularity. The former refers to the assumption that the successful acquisition of language is assumed to be more likely to take place when the target structure

occurs in an anchor position whereas the latter refers to the notion that regular structures are easier to acquire than the opposite. In this case, the TQ is perceptually fairly salient, as it occurs at the end of the sentence. In addition, the TQ applies to the general word order used to form yes-no questions in English, and accordingly, the TQ meets the regularity criteria.

In contrast to the TQ rule, the NA and CON structures reflected a distinct effect in improvement compared to the TQ across the sample and in each modality group in Experiment 1, and only the CON structure in the visual group for Experiment 2. This marked improvement occurs although three factors contributing to conceptual complexity are in operation in the rule system governing them: specifically, its perceptual non-salience, irregularity, and functional complexity, all of which produce processing difficulties. Although the TQ and CON structures share a basic syntax system (i.e., both involve two clauses), the former is assumed to place higher demands on linguistic processing due to its inclusion of complex functions. That is, the TQ syntax involves a declarative (or an imperative) clause in which an L2 learner should refer back to the verb phrase to generate a syntactically correct auxiliary verb involved in forming a TQ structure. This is relatively unlike the CON structure, where an L2 learner does not need to refer back to the verb phrase of the conditional clause to build up the dependent clause; instead, the dependent clause has a fixed structure, which can play a facilitative role in its learning (i.e., would + have + past participle). The difference between the two structures in terms of complexity in function could probably contribute to the differential improvement in performance between the two (see the literature review under Section 3.3 for more discussion).

The lack of a distinct improvement in performance of the NA structure compared to the TQ in Experiment 2 might be explained by participants' limited available automated knowledge of their English syntax prior to exposure to the interventions. This means participants already had automatized knowledge of the English word order SVO, therefore,

abundant cognitive effort is required for the L2 learner's grammatical knowledge to be re-automatized such that he or she can generate the abnormality of the word order (i.e., an auxiliary verb-subject inversion) without conscious consideration or with only limited consideration. In addition, the TQ structure obtained slightly better improvement in Experiment 2 than Experiment 1 and consequently this slight improvement probably played a role beyond the lack of the NA to elicit a higher significant improvement than the TQ in Experiment 2.

The significant variation in the amount of improvement across syntactic structures, as determined by EIT, appears to not be influenced by learners' L1. Both experiments showed a significantly poorer improvement in the TQ performance than in the CON although there is an apparent difference in the TQ syntax between Arabic and Chinese. Arabic has only a fixed form of the question tag *isn't it so?*, while Chinese can freely possess three tag patterns — verb-not-verb, verb-particle, and negative-verb-particle — where each pattern functions for distinct discourse purposes. Although the comparisons between English and Chinese/ English and Arabic revealed that the CON structure is difficult to learn, it reflected a distinct improvement compared to the TQ in both experiments. Finally, the NA structure reflected a distinct improvement compared to the TQ in Experiment 1, but this distinct effect was absent in Experiment 2 although both languages (Arabic and Chinese) have a similar level of difficulty in terms of the NA structure (note: further research and analysis is necessary to verify these conclusions).

The NA structure, for instance, reflected a distinct improvement compared to the TQ in Experiment 1, but this distinct effect was absent in Experiment 2 although both languages (Arabic and Chinese) have a similar level of difficulty in terms of the NA structure. That is, both languages have a number of lexical words expressing NA, and these adverbs do not cause subject-verb inversion when the adverb occurs at the beginning of a sentence as they do

in English. For the TQ and CON structures, both experiments showed a significantly poorer improvement in the TQ performance than in the CON although there is an apparent difference in the TQ syntax between Arabic and Chinese. Arabic has only a fixed form of the question tag *isn't it so?*, while Chinese can freely possess three tag patterns — verb-not-verb, verb-particle, and negative-verb-particle — where each pattern functions for distinct discourse purposes. The shared property in the TQ between the two languages is that the positivity or negativity of the statement is irrelevant to the tag question itself. Finally, for the CON structure, there is also a difference between Arabic and Chinese. In the former, the CON is encoded by the conjunction *Law* whereas in the latter, it is determined by the context, time reference (such as 'last week' or 'now'), and internal semantic logic using one verb-tense form.

Two conclusions can be drawn from the above-mentioned interpretations related to syntactic structures. First, the variation in automating the L2 knowledge of the three syntactic structures (TQ, NA, and CON) was determined by the EIT, which could provide critical data on the constraints of the developmental process of distinct syntactic structures. In particular, performance in the EIT hinges on fluent processing in the auditory input and oral output as a dynamic, real-time entity, and consequently, a lack of automatic processing for either of the two will result in performance detriments. This is unlike the GJT, which primarily requires only fluent visual recognition of linguistic forms.

Second, more critical data can be elicited from an empirical study when the stimuli are infrequent and uncommon to L2 learners. This is manifested in the NA and CON structures, both of which are limited in language use and typically are used in more formal written English (as affirmed by an analysis of corpora in conditionals and an analysis of the British National Corpus). This can be evidenced by the poor performance on the pretest in both experiments. This is in contrast to the TQ structure, which is frequently used in daily

communication (based on the British component of the International Corpus of English). Hence, it appears to be common to L2 learners. This can be supported by the high mean score in the pretest performance, which, in turn, may play a role in supressing any significant improvement.

The results pertaining to the second research question, whether WM plays a differential role in the successful acquisition of implicit grammatical knowledge under exposure to one of the two modalities, are in line with the expectations of the findings of the first research question in each experiment. In Experiment 1, the result that visual intervention clearly tended to be more effective than auditory intervention at fostering L2 automatic processing, as determined by both the GJT and the EIT, was supported by the significant relationship between WM capacity and the GJT. That is, the auditory intervention imposed greater computational demands on WM resources, leading to detrimental effects on language development. In Experiment 2, a similar pattern of improvement was obtained between the two modality interventions, and consequently, WM capacity did not mediate the effects of the auditory or visual modalities on language development. An initial conclusion can be drawn that the more the modality intervention (i.e., visual modality) induced linguistic knowledge to become available for use in automatic processing, the less were processing demands exercised on the WM resources, and vice versa. Moreover, in both experiments, phonological short-term memory (as measured by a digit span task) was found not to correlate with the EIT performance, suggesting the capacity for storing information in short-term memory did not greatly affect EIT performance.

6.4   Reflection on EIT and GJT

Although responses to both the timed GJT and the EIT reflect underlying automatic processing in a similar way, it can be observed that performance on the EIT was considerably poorer than performance reflected on the GJT in both testing sessions as shown in Table 52

below. This poor performance reflected on the EIT in comparison to performance on the GJT

is also observed in previous validation studies (e.g., Bowles, 2011; R. Ellis, 2005; J.-e. Kim

& Nam, 2017; Spada et al., 2015; Zhang, 2015).

Table 52

Differences in Mean (M) Raw Scores and Standard Errors (ES) between Timed GJT and EIT

Condition	Timed GJT		EIT	
	M	*SE*	*M*	*SE*
Experiment 1				
Pretest	12.49	0.37	6.18	0.33
Posttest	16.57	0.38	8.95	0.46
Mean change	4.07	0.45	2.76	0.33
Experiment 2				
Pretest	6.84	0.44	5.10	0.44
Posttest	12.55	0.54	7.47	0.61
Mean change	2.71	0.51	2.36	0.47

The most likely explanation for poor performance on the EIT may pertain to the

underlying variables operationalised by the EIT. Good performance on this type of measure,

in contrast to the GJT, entails high fluency in auditory input and oral output, both of which

hinge on automatized processing. Put differently, learners with poor auditory abilities spend

more processing time on individual items due to a lack of automaticity in aural recognition

(Ringbom, 2007). In contrast, in the case of the GJT, possession of a fairly broad lexical

knowledge of L2 English might be sufficient to achieve a reasonable level of reading

competence. In the EIT, strength in other core abilities (e.g., oral repertoires, variation in

prosodic signals and on-line processing, or achievement of a complementary relationship

between top-down and bottom-up processing) is required in order to process the rapid,

auditorily presented linguistic stream, which is an essential ability for reasonable oral

language production as required in the ELT.

Another possible explanation relates to Segalowitz's (2003, p. 400) claim that "if one can handle the phonology and syntax of a second language automatically, then more attention can be paid to processing semantic, pragmatic, and sociolinguistic levels of communication". This implies, in the case of the EIT, that if a participant is asked to allocate his/her attention to meaning while his or her L2 phonology and syntax is still not automatized to a sufficient extent, the participant's comprehension of aurally presented speech is likely to be detrimentally affected to a greater extent than in the opposite case; conversely, if the L2 phonology and syntax is sufficiently automatized, some (although limited) attentional resources may be left over, allowing meaning to be processed.

Another significant factor is that despite the time pressure exerted by the GJT, there may be still some opportunity for the participant to engage in repeated processing and monitoring. This is in contrast to the EIT, in which the participant has no control over the rate at which the aural stream is presented because such processing requires the continuous aural stream to be segmented and comprehended while new material is being perceived.

Moreover, another crucial factor that potentially accounts for the low performance on the EIT compared to the GJT relates to the difference in the scoring criteria between the two measures. In the GJT, participants were required to select one of three options: 'correct', 'incorrect', or 'not sure'. Responses were scored dichotomously as correct (one point) or incorrect (zero points), with items to which the 'not sure' response or no response was given being treated as incorrect. Scoring of responses to the EIT, in contrast, requires human raters to assess oral responses, which makes it challenging to ensure consistent application of the scoring criteria in cases of partially correct responses. This issue can be illustrated by a response like *Seldom does we get the chance to watch television together.* In this instance, the participant was able to move the auxiliary verb 'does' to its appropriate position, but the subject and verb did not agree with one another. Another issue can be reflected in a response

like *Had the storm continued, we would have loss the house.* In this instance, it is unclear if the participant used the past participle of *lose* in a pronunciation like /lɒst/ or in a pronunciation without /t/. Although responses were scored dichotomously for the production of correct or incorrect English, the scoring criteria are not as straightforward and objective as those used for the timed GJT.

Given that the judging criteria for the EIT is a blunt instrument and cannot be constrained in the same way as the criteria used in the GJT, several criteria were cautiously and meticulously established to enable the raters to score responses consistently. This difference in the judging criteria between the EIT and the timed GJT which suggests that the potential for a response to be scored incorrect is greater in the former case than in the latter.

Lastly, and more importantly, the fact that elements of the inherent nature of the EIT most probably played a role in the low performance of the participants does not imply that the accuracy of this measure is poor; instead, these elements of the EIT are largely considered to be unique idiosyncrasies that enhance its validity, provided that any given version of the EIT is carefully designed and pilot validation studies are conducted. Ellis and Rover (2018) concluded, based on previous factor-analysis studies, that the EIT is the most robust available measure of implicit grammatical knowledge.

Chapter 7: Conclusion

This study investigates the (auditory vs. visual) modality effects on L2 implicit knowledge acquisition and how WM capacity interacts with the acquisition of this knowledge accrued in each input modality. The results of the current research demonstrated no difference between the effects of auditory and visual input modalities on automatizing L2 grammatical knowledge. However, as there was an observed trend towards the visual modality advantage in the acquisition of implicit knowledge – as determined by both measures: the EIT and the timed GJT – and there was an interaction between WM capacity and the development of L2 implicit knowledge accrued from the auditory modality, one can conclude from the results that modality effects is another relevant factor in L2 acquisition, which might indeed play a differential role in stimulating the L2 automatic processing. Furthermore, the results suggest that research which examines modality effects on L2 acquisition, learning, or performance, without taking the role of WM capacity into account (i.e., using a measure of WM capacity), may not obtain a complete picture of modality effects.

Although some researchers (e.g., Penney, 1989; Murphy, 1997) have argued that it would be unreasonable to suggest that one modality is a superior avenue to underlying grammatical competence than the other, in this experiment, a tendency towards the visual modality advantage in language acquisition was found in both the timed GJT and the EIT data; moreover, this advantage was not influenced by the variation in performance of syntactic structures (TQ, NA and CON) between the auditory and the visual condition. One can be more confident, thus, in suggesting that modality may indeed affect the L2 acquisitional process. That is, following an emergent view, once a link is established between a particular form and its meaning, that link tends to be more consistently strengthened by the visual input modality than the auditory modality.

Importantly, however, prior to generalising the results of the present study or drawing any definitive conclusions, a replication of this study must be carried out, with several modifications.

7.1   Limitations

First, the main source of limitations (i.e., error) arises from the time limit established for the timed GJT's version A and B. In particular, the time limit for each sentence in version A and B was established based on native speakers' average response time. However, the error that occurred was that the same time limit established for each sentence in the version A was inadvertently used for version B. However, significant efforts were made to make the two versions similar in terms of the length of sentences, structures, and difficulty. In addition, it should be noted that the same participant in each experiment performed both versions. That is, if a participant was pretested using version A, he/she would be posttested using version B and vice versa.

An independent sample t-test was performed on the pretest/posttest scores to rule out differences between versions A and B. The analysis revealed no difference between versions A and B on the pretest scores in Experiment 1 with Chinese participants, t (75) = 0.50, p = 0.61, but a difference was observed between the two versions in the posttest scores t (75) = - 2. 45, p = 0.01. For experiment 2 with Arabic participants, no significant difference was observed between the two versions in the pretest scores t (35) = 0.77, p = 0.44 or posttest scores t (35) = 0.118, p = 0.90, suggesting that this error should not be much a cause for concern.

Second, although the number of Arabic participants in Experiment 2 is higher than in several previous studies (e.g., R. Ellis et. al, 2009; Reinders & R. Ellis, 2009), a sample of 30 participants and above would be optimal, specifically for ANOVA tests. Due to ecological

considerations (i.e., limited access to Arabic speakers in Bristol) and the time allocated for this work to be completed, a smaller sample was aimed at.

Third, although the timed GJT was designed in accordance with the guidelines established by previous validation studies (e.g., Ellis, 2006; Loewen, 2009) and the time limit for the response to each sentence established based on L1 speakers' performance, the additional 20% of the time limit of each sentence was added to each sentence in the current study for L2 participants. This choice might cause some concern because it is not known how much extra time L2 readers need to judge compared to L1 readers. A similar concern is also applied to the threshold for response times in the EIT. Although most previous validation studies did not establish a fixed time limit for the latency between prompt and response, and pressure time was left to the discretion of the administrator running the task, the time limit in this study was established based on the length of an utterance, and then 3 seconds were added to the length of each utterance. However, this time limit was established without an empirical reason and so is a cause for some concern.

Fourth, although the present work aimed to examine the successful emergence of L2 implicit knowledge, further reliable data regarding the measures of implicit knowledge would be obtained if explicit knowledge measures (e.g., untimed GJT, metalinguistic knowledge test, see R. Ellis, 2005) were employed. This will offer more insights into whether the timed GJT and the EIT (as measures of implicit knowledge) and the untimed GJT and metalinguistic knowledge test (as measures of explicit knowledge) can tap into two distinct types of knowledge.

7.2   Further research

Two lines of further investigation have been extended from the present work. The first is concerned with the impact of the interaction between the type of learning mechanism and modality effects on L2 implicit knowledge acquisition. That is, it is interesting to examine

how modality effects are manifested under different learning conditions, such as only input, input alongside output, incidental learning, enhancement input, and explicit learning (see Sanz et al., 2014, for a comprehensive review). For instance, in the case of explicit and implicit learning, the more the nature of the learning mechanism requires explicit learning, the more controlled processing (i.e., large amount of attentional resources) is involved, which, in turn, makes greater demands on WM capacity (Y. Wang, 2017, see Tagarelli et al., 2015 for empirical evidence). This probably suggests that the heavy processing demands accruing from explicit learning will be more likely to be manifest in the processing in the auditory modality than in the visual modality. This is because explicit learning requires more attentional resources than implicit learning, and at the same time, the processor must parse the rapid auditory input (including acoustic and linguistic elements) resulting in more factors competing for the learner's limited WM resources. Hence, further research is needed to verify these hypotheses.

The second research implication is related to the interaction between modality effects on the successful emergence of L2 implicit knowledge and learners' developmental readiness. Research on the acquisition of grammatical structures (e.g., question formation, past tense, and negation) has been shown to proceed through a fixed sequence of stages in that learners can only progress to the next stage once the earlier stages have been successfully acquired (Loewen & Reinders, 2011). Ellis and Shintani (2013, p. 74) claimed that "if learners are not ready to acquire a specific feature, they will not acquire it no matter how intensively it is taught". An example of the developmental sequence is shown in the English irregular past tense, as provided by Loewen and Reinders (2011). L2 learners first use (e.g., he went) it correctly because they learn the form as an individual lexical item. Then they learn that the morpheme *-ed* is added to the end of the verb to form the regular past tense. Consequently, then they apply this rule to irregular verbs (i.e., overgeneralise) they need to

use prior to successfully acquiring regular and irregular past tense. Finally, learners fully acquire regular and irregular past tense, and hence they use the correct irregular forms (i.e., he ate). Thus, research into developmental readiness claims that L2 learners' developmental readiness in language learning should be taken into consideration to demonstrate the linguistic input supplied and the learners' readiness to acquire it coincide. Li (2015) suggested it is necessary to investigate the role of proficiency in mediating the impact of L2 learners' developmental readiness on the acquisition of a certain grammatical structure. Therefore, further work could provide clearer evidence on the extent to which the learners' readiness to acquire the L2 implicit knowledge of a given structure is affected by the modality (auditory vs. visual) in which this structure is presented. For experiments of this type, participants, for example, can be placed into three proficiency groups (e.g., beginning, intermediate, and advanced levels) based on their pretest performance. Such a method can further open a more apparent window on the role of WM capacity in the different developmental stages of L2 learners' implicit grammatical knowledge.

7.3   Recapitulation

The overall purpose of the study was to determine i) whether encoding modality (auditory vs visual) affects L2 implicit grammatical knowledge acquisition differently, and ii) whether WM capacity predicts how this implicit knowledge is acquired differently depending the encoding modality. The results of the two experiments demonstrated no significantly differentiable effects between the visual and the auditory interventions in automatizing L2 grammatical knowledge, but better automaticity development was clearly observed in the visual intervention in only Experiment 1, as determined by both tasks: the timed GJT and the EIT. WM capacity was found, in Experiment 1 only, to predict the L2 automaticity development accrued from auditory exposure in the GJT, but not in the EIT. I concluded that

modality effects as another relevant factor of L2 acquisition should be taken into account when investigating L2 implicit knowledge acquisition in future research.

Notes

1. One participant in Experiment 1 received 7.0 in IELTS, but the participant was not removed from the analysis because, for some reason, the participant was not recognized.

2. One participant in Experiment 2 was posttested 12 days following the first session.

3. The digit span, which was only used to control EIT, appears to have a minor fault. In this task, a participant was required to memorize a series of digits they heard and was asked to recall the digits in the order presented by typing them. However, it was noted by one participant from Experiment 1 at the beginning of the data collection stage that the task could record the digits the participant typed even if the participant was still hearing the digits. Therefore, one more sentence was added to the instruction: 'REMEMBER that DO NOT start typing the digits until the task asks you to type'. Because the digit span is not the main focus of the study, the participant was not eliminated from the whole study, but the participant was asked to repeat the task.

4. Nearly two participants from Experiment 2 who had the timed GJT crashed at the very beginning of the test.

5. Nearly three participants from Experiment 2, at the beginning of data collection stage, were only exposed to three sentences in the practice stage of the timed GJT, but the task had 6 sentences. Together with point 3, these cases do not appear to affect the results enough that those participants to be eliminated from the whole study, particularly there is limited access to Arabic participants in Bristol.

6. Nearly 4 participants who had the OSPAN crashed during the instruction section, which includes practice as well, before they started the real test. Such this task is complicated

and lasts around half an hour. Therefore, the possibility of having effects on the results is very low, if no effects at all.

7. One participant completed the OSPAN twice, but the first score was unavailable. Because the OSPAN does not measure a type of learning condition and sufficient time (one month and a half) passed between the first and second testing to avoid any previous practice effects, no concern was associated with using the second score.

8. Age information for one participant in each group in Experiment 1 is unavailable. The same information is unavailable for one participant in the visual group in Experiment 2.

References

Abu-Shquier, M., & Abu Shqeer, O. (2012). Words ordering and corresponding verb-subject agreements in English-Arabic machine translation: Hybrid-based approach. *International Journal of Soft Computing and Software Engineering, 2*, 49-60. doi:10.7321/jscse.v2.n8.5

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological bulletin, 135*(1), 50-68. doi:10.1037/a0014411

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological bulletin, 131*(1), 30.

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research, 16*(1), 9-37.

Alderson, J. C. (2005a). *Diagnosing foreign language proficiency : the interface between learning and assessment*. London: Continuum.

Alderson, J. C. (2005b). *Diagnosing foreign language proficiency: The interface between learning and assessment*: A&C Black.

Alloway, T. P. (2010). *Improving working memory: Supporting students' learning*. London: Sage.

Alptekin, C., & Ercetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*(2), 206-219. doi:10.1111/j.1467-9817.2009.01412.x

Anderson, J. (1980). *Cognitive psychology and its implications*. San Francisco: W. H. Freeman.

Anderson, J. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.

Anderson, J. (1985). *Cognitive psychology and its implications* (2th ed.). New York: Freeman.

Andrade, J. (2002). *Working memory in perspective*. Hove: Taylor & Francis.

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*, 49-78. doi:10.1111/j.1467-9922.2012.00706.x

Arc-communications. (2011). The difference between written and spoken Chinese. Retrieved from https://www.arc-japanese-translation.com/chinese/03tidbits.html

Avons, S. E., & Phillips, W. A. (1980). Visualization and memorization as a function of display time and poststimulus processing time. *Journal of Experimental Psychology: Human Learning and Memory, 6*(4), 407.

Ayoun, D. (2005). *Parameter setting in language acquisition*. London: A&C Black.

Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 54-73). New York: Cambridge University Press.

Baddeley, A. D. (1992). Working memory: The interface between memory and cognition. *Journal of cognitive neuroscience, 4*(3), 281-288.

Baddeley, A. D. (1999). *Essentials of human memory*. Hove: Psychology Press.

Baddeley, A. D. (2002). *Human memory: Theory and practice*. Hove: Psychology Press.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208. doi:https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D. (2007). Working memory: Multiple models, multiple mechanisms. *Science of memory: Concepts*, 151-153.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological review, 105*, 158-173. doi:10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working memory. In A. D. Baddeley, G. Hitch, & G. Bower (Eds.), *The psychology of learning and motivation* (pp. 47–90). London: Academic.

Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research, 5*(2), 95-127.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73-83. doi:10.1016/j.intell.2014.05.007

Berent, G. P. (1985). Markedness considerations in the acquisition of conditional sentences. *Language Learning, 35*(3), 337-372. doi:10.1111/j.1467-1770.1985.tb01082.x

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge: Cambridge University Press.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245-261). New Jersery: Lawrence Erlbaum.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 33*(02), 247-271.

Broadbent, D., Vines, R., & Broadbent, M. (1978). Recency effects in memory, as a function of modality of intervening events. *Psychological Research-Psychologische Forschung, 40*(1), 5-13. doi:10.1007/bf00308459

Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly, 49*(1), 141-168. doi:10.1002/tesq.168

Buck, G. (1997). The testing of listening in a second language. In C. M. Clapham & D. Corson (Eds.), *Language Testing and Assessment* (Vol. 7, pp. 65–74). Dordrecht: Encyclopedia of Language and Education.

Bygate, M. (2001). Speaking. In R. Carter & D. Nunan (Eds.), *The CambridgeGuide to Teaching English to Speakers of Other Languages* (pp. 14-20). Cambridge: Cambridge University Press.

Cardierno, T. (1995). Formal instruction from a processing perspective - an investigation into the Spanish past tense. *Modern Language Journal, 79*(2), 179-193.

Chai, J., & Erlam, R. (2008). The Effect and the Influence of the Use of Video and Captions on Second Language Learning. *14*, 25-44.

Chappell, H., & Lan, L. (2016). The Routledge encyclopedia of the Chinese language. In S.-w. Chan (Ed.), *Mandarin and other sinitic languages* (pp. 605-628). New York: Routledge.

Chaudron, C. (2003). Data collection in SLA research. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 762-828). Malden: Blackwell Publishing.

Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious structural knowledge of form-meaning connections. *Consciousness and Cognition, 20*(4), 1751-1760. doi:10.1016/j.concog.2011.03.003

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, Mass ; London: MIT Press.

Chomsky, N. (1986). *Knowledge of language : its nature, origin, and use*. New York ; London: Praeger.

Chou, C.-L. (2000). Chinese speakers' acquisition of English conditionals: Acquisition order and L1 transfer effects. *Second Language Studies, 19*.

Chun, D. M., & Payne, J. S. (2004). What makes students click: Working memory and look-up behavior. *System, 32*(4), 481-503. doi:https://doi.org/10.1016/j.system.2004.09.008

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences, 2*(10), 406-416. doi:https://doi.org/10.1016/S1364-6613(98)01232-7

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36*(6), 584-606.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5*(6), 365-373. doi:10.3758/BF03210660

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi:https://doi.org/10.1016/S0022-5371(80)90312-6

Daniel, D. B., & Woody, W. D. (2010). They hear, but do Not listen: retention for podcasted material in a classroom context. *Teaching of Psychology, 37*(3), 199-203. doi:10.1080/00986283.2010.488542

Danks, J. (1990). Comprehension in listening and reading: Same or different? In J. Danks & K. Pezdek (Eds.), *Reading and understanding* (pp. 1-39). Newark, DE.

Davies, W. D., & Kaplan, T. I. (1998). Native speaker vs. L2 learner grammaticality judgements. *Applied Linguistics, 19*(2), 183-203. doi:10.1093/applin/19.2.183

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*(3), 529-555. doi:10.1111/j.1467-1770.1996.tb01246.x

Decoo, W. (1996). The induction-deduction opposition: Ambiguities and complexities of the didactic reality. *Iral-International Review of Applied Linguistics in Language Teaching, 34*(2), 95-118. doi:10.1515/iral.1996.34.2.95

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313-348). Malden: Blackwell.

Dennis, I. (1977). Component problems in dichotic listening. *The Quarterly Journal of Experimental Psychology, 29*(3), 437-450.

Detey, S. (2005). *Interphonology and orthographic representations. On the role of written input in the teaching / learning of oral French by Japanese learners (Unpublished doctoral dissertation).* Université Toulouse le Mirail, France.

Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). Cambridge: CUP.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 206– 257). New York, NJ: Blackwell.

Doughty, C. J. (2004). Effects of instruction on learning a second language: A critique of instructed SLA research. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 181-202). Mahwah: Lawrence Erlbaum Mahwah.

Downing, A., & Locke, P. (2006). *English grammar: A university course*. Abingdon: Routledge.

Dussias, P. (2003). Syntactic Ambiguity Resolution in L2 Learners: Some Effects of Bilinguality on L1 and L2 Processing Strategies. *Studies in Second Language Acquisition, 25*, 529-557.

Ellis, N. (1994). Implicit and explicit language learning. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79-114). San Diego: Academic Press.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition, 24*(2).

Ellis, N. (2003). Constructions, chunking and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell Publishing.

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning* (pp. 35-47). Washington: Georgetown University Press.

Ellis, R. (1991). Grammatically judgments and second Language acquisition. *Studies in Second Language Acquisition, 13*(02), 161-186.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(02), 141-172.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *Tesol Quarterly, 40*(1), 83-107. doi:10.2307/40264512

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 1-25). Bristol: Multilingual Matters.

Ellis, R. (2009b). Investigating learning difficulty in terms of implicit and explicit knolwedge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 143-166). Bristol: Multilingual Matters.

Ellis, R. (2015a). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 417-441). Amsterdam: John Benjamins.

Ellis, R. (2015b). *Understanding second language acquisition* (Second edition. ed.). Oxford: Oxford University Press.

Ellis, R., & Roever, C. (2018). The measurement of implicit and explicit knowledge. *The Language Learning Journal*, 1-16. doi:10.1080/09571736.2018.1504229

Ellis, R., Shawn, L., & Erlam, R. (2009). Implicit and explicit corrective feedback and the acquisition of L2 grammar. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H.

Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 303-332). Bristol: Multilingual Matters.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research, 9*(2), 147-171.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An Empirical validation study. *Applied Linguistics, 27*(3), 464-491.

Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 464–491). Bristol: Multilingual Matters.

Erlam, R., Loewen, S., & Philp, J. (2009). The roles of output-based and input-based instruction in the acquisition of L2 implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 241-261).

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams. In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the 2010 Second Language Research Forum* (pp. 18-28). Somerville, MA: Cascadilla Proceedings Project.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT journal, 57*(4), 325-334.

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics, 17*(3), 319-331.

Fulcher, G. (1991). Conditionals revisited. *ELT Journal, 45*(2), 164-168.

Gass, S. M. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gass, S. M. (2018). *Input, interaction, and the second language learnerRoutledge linguistics classics* (pp. 1 online resource).

Gass, S. M., & Selinker, L. (2001). *Second language acquisition : an introductory course* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition : an introductory course* (3rd ed.). New York ; London: Routledge.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543. doi:10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Hove: Psychology Press.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*(1), 65-77.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*(2), 103-127.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76. doi:https://doi.org/10.1016/S0010-0277(98)00034-1

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75. doi:https://doi.org/10.1016/S0346-251X(99)00060-3

Goldstein, E. (2014). *Cognitive psychology: Connecting mind, research and everyday experience*. Stamford: Cengage Learning.

Gower, R., & Walters, S. (1995). *Teaching practice handbook: A reference book for EFL teachers in training*. New York: Heinemann.

Gravetter, F., & Forzano, L.-A. (2015). *Research methods for the behavioral sciences*. New York: Nelson Education.

Gutierrez, X. (2013). The construct validity of grammaticality judgement tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition, 35*(3), 423-449. doi:10.1017/s0272263113000041

Haberlandt, K. (1997). *Cognitive psychology*: Allyn & Bacon.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited extending williams (2005). *Studies in Second Language Acquisition, 32*(3), 465-491. doi:10.1017/s0272263110000045

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in second language acquisition, 14*(1), 25-38.

Hartsuiker, R., & Barkhuysen, P. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes, 21*, 181-204. doi:10.1080/01690960400002117

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington: Georgetown University Press.

Hsin, A. L. C. (2016). An analysis of Chinese tag questions with a cross-linguistic comparison to English tags. *14*, 69-119. doi:10.6519/TJL.2016.14(1).3

Hsu, W. (2017). The modality effect of input repetition on vocabulary consolidation. *Konińskie Studia Językowe, 5*(4), 403-424.

Hulstijn, J. (2008). Incidental and intentional learning. In M. Long & C. J. Doughty (Eds.), *The handbook of second language acquisition* (pp. 349-381). Malden: Blackwell.

Hulstijn, J. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-47). Amsterdam: John Benjamins Publishing Company.

Ioup, G. (1995). Evaluating the need for input enhancement in post-critical period language acquisition. In D. M. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition: A critical look at the critical period hypothesis* (pp. 95-123). Clevedon England: Philadelphia: Multilingual Matters.

Izumi, S. (2002). Output, input enhancement and the noticing hypothesis. *Studies in Second Language Acquisition, 24*(4), 541-577.

Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition, 12*(1), 65-82. doi:10.1017/S1366728908003908

Jackson, S. L. (2014). *Research methods: A modular approach*. Stamford: Cengage Learning.

Jensema, C., Danturthi, R., & Burch, R. (2001). Time Spent Viewing Captions on Television Programs. *American annals of the deaf, 145*, 464-468. doi:10.1353/aad.2012.0144

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*. doi:10.1111/lang.12034

Jiang, W. (2009). *Acquisition of word order in Chinese as a foreign language* (Vol. 38). New York: Walter de Gruyter.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language learning, 42*(2), 217-248.

Joseph, W. A. (2019). Politics in China: an introduction. In W. A. Joseph (Ed.), *Studying Chinese Politics* (pp. 3-40). New York: Oxford University Press.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 199-225. doi:10.1111/j.0079-1636.2004.00135.x

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review, 99*(1), 122-149. doi:10.1037/0033-295x.99.1.122

Kebbe, M., & Kebbe, M. Z. (2000). *A transformational grammar of modern literary Arabic*. London: Routledge.

Kerz, E., Wiechmann, D., & Riedel, F. B. (2017). Implicit learning in the crowd investigatingthe role of awareness in the acquisition of l2 knowledge. *Studies in Second Language Acquisition, 39*(4), 711-734. doi:10.1017/s027226311700002x

Kim, J.-B., & Ann, J.-Y. (2011). English Tag Questions: Corpus Findings and Theoretical Implications. *English Language and Linguistics, null*. doi:10.17960/ell.2008..25.006

Kim, J.-e., & Nam, H. (2017). Measures of implicit knowledge revisited processing modes, time pressure, and modality. *Studies in Second Language Acquisition, 39*(3), 431-457. doi:10.1017/s0272263115000510

Kirkpatrick, A. (1993). Information sequencing in modern standard Chinese. *Australian Review of Applied Linguistics, 16*(2), 27-60.

Kormos, J., & Safar, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism-Language and Cognition, 11*(2), 261-271. doi:10.1017/s1366728908003416

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st ed.). Oxford ; New York: Pergamon.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading - additional evidence for the input hypothesis. *Modern Language Journal, 73*(4), 440-464. doi:10.2307/326879

Krashen, S. D. (1998). Comprehensive output? *System, 26*(2), 175-182.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.

LaRue, A. (2013). *Aging and neuropsychological assessment*. New York: Springer Science & Business Media.

Lawson, R., Fernandes, A. M., Albuquerque, P. B., & Lacey, S. (2015). Remembering touch: Using interference tasks to study tactile and haptic memory. In P. Jolicoeur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory* (pp. 239-259). San Diego: Academic Press.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Secomnd Language Acquisition, 25*, 37-63. doi:10.1017/S0272263103000020

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229-270. doi:10.1111/j.1467-9922.2007.00408.x

Levelt, W. J. M. (1989). *Speaking : from intention to articulation*. Cambridge, Mass.: MIT Press.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, C. N., & Thompson, S. A. (2009). The world's major languages. In B. Comrie (Ed.), *The world's major languages* (pp. 703-723). Abingdon: Routledge.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94-112). Bristol: Multilingual Matter.

Loewen, S. (2014). *Introduction to instructed second language acquisition*. New York: Routledge.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamin.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Wlliams (Eds.), *Focus on Form in Classroom Language Acquisition* (pp. 15-41). Cambridge: CUP.

Lu, J. H. T. (1980). A Study of quantifiers in Mandarin Chinese. *Journal of the Chinese Language Teachers Association, 15*(3), 1-23.

Mackey, A., & Gass, S. M. (2005). *Second language research : methodology and design*. Mahwah ; London: Erlbaum.

Mackey, A., & Philp, J. (1998). Conversational Interaction and Second Language Development: Recasts, Responses, and Red Herrings? *The Modern Language Journal, 82*(3), 338-356. doi:10.1111/j.1540-4781.1998.tb01211.x

MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition, 19*(02), 277-281.

Martin, K. I., & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379-413. doi:10.1017/S0272263112000125

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

Mazur, J. E. (2015). *Learning and behavior: Instructor's review copy*: Psychology Press.

McGinnis, S. (1988). Sentential Word Order in Chinese: Towards a Theoretical/Pedagogical Interface. *Journal of the Chinese Language Teachers Association*.

McLaughlin, B. (1987). *Theories of second-language learning*. London: Edward Arnold.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11*(2), 113-128. doi:10.1093/applin/11.2.113

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: an information-processing perspective. *Language Learning, 33*(2), 135-158. doi:10.1111/j.1467-1770.1983.tb00532.x

McLoughlin, L. (2002). *Colloquial Arabic (Levantine)*. London: Routledge.

Meisel, J. M. (2011). *First and second language acquisition: Parallels and differences*. Cambridge: Cambridge University Press.

Mertler, C. A. (2019). *Introduction to educational research* (Second edition. ed.). Los Angeles: SAGE Publications, Inc.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). New Jersery: Lawrence Erlbaum.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A.

Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442-481). Cambridge: Cambridge University Press.

Montrul, S. (2005). On knowledge and development of unaccusativity in Spanish L2 acquisition. *Linguistics, 43*(6), 1153-1190.

Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition, 28*(1), 31-65. doi:10.1017/s0272263106060025

Morgan-Short, K., Marsden, E., Heil, J., Issa, B. I., Leow, R. P., Mikhaylova, A., . . . Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and reading comprehension (vol 68, pg 92, 2018). *Language Learning, 68*(3), 851-851. doi:10.1111/lang.12308

Muijs, D. (2010). *Doing Quantitative Research in Education with SPSS*: SAGE Publications.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 227-243). New Jersey: Lawrence Erlbaum.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). Cambridge: Cambridge university press.

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research, 13*(1), 34-65.

Mystkowska-Wiertelak, A., & Pawlak, M. (2012). *Production-oriented and comprehension-based grammar teaching in the foreign language classroom*. Berlin ; New York: Springer.

Nabei, T., & Swain, M. (2002). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness, 11*(1), 43-63.

Nassaji, H. (2013). Participation Structure and Incidental Focus on Form in Adult ESL Classrooms. *Language Learning, 63*(4), 835-869. doi:10.1111/lang.12020

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528. doi:10.1111/0023-8333.00136

Norris, R. W. (2003). How do we overcome the difficulties of teaching conditionals? *Bulletin of Fukuoka International University, 9*, 39-50.

Nursey, J., & Phelps, A. J. (2016). Stress, trauma, and memory in PTSD. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior* (pp. 169-176). San Diego: Academic Press.

Osaka, M., & Osaka, N. (2007). Neural bases of focusing attention in working memory: An fMRI study based on individual differences. In M. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 99-118). New York: Oxford University Press.

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31*(2), 117-118. doi:10.3758/BF03334156

Papagno, C., Lucchelli, F., & Vallar, G. (2008). Phonological recoding, visual short-term store and the effect of unattended speech: Evidence from a case of slowly progressive anarthria. *Cortex, 44*(3), 312-324. doi:https://doi.org/10.1016/j.cortex.2006.10.001

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (Vol. 393, pp. 393–419). London: Academic Press.

Park, G. P. (2004). Comparison of L2 listening and reading comprehension by university students learning English in Korea. *Foreign Language Annals, 37*(3), 448-458. doi:10.1111/j.1944-9720.2004.tb02702.x

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*(4), 398-422. doi:10.3758/bf03202613

Pereltsvaig, A. (2017). *Languages of the world: An introduction*. Cambridge: Cambridge University Press.

Pervin, L. A., & John, O. P. (1999). *Handbook of personality: Theory and research*. New York: Elsevier.

Pienemann, M. (1998). *Language processing and second language development: Processability theory* (Vol. 15). Amsterdam: John Benjamins.

Pietrzykowska, A. (2011). The influence of visual input enhancement on the acquisition of English embedded questions. In M. Pawlak (Ed.), *Extending the boundaries of research on second language learning and teaching* (pp. 3-13). Heidelberg: Springer.

Prutting, C. A., Gallagher, T. M., & Mulac, A. (1975). Expressive portion of NSST compared to a spontaneous language sample. *Journal of Speech and Hearing Disorders, 40*(1), 40-48. doi:10.1044/jshd.4001.40

Pujolǎ, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL, 14*(2), 235-262. doi:10.1017/S0958344002000423

Randall, M. (2007). *Memory, psychology and second language learning*. Amsterdam: John Benjamins.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(04), 829-856.

Rebuschat, P., & Williams, J. N. (2013). Implicit learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 20). Oxford: Blackwell Publishing Ltd.

Reinders, H. (2010). *The effects of task type and instructions on second language acquisition*. Newcastle: Cambridge Scholars.

Reinders, H., & Ellis, R. (2009). The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 281-302). Bristol: Multilingual Matters.

Richards, J. C. (2015). *Key issues in language teaching*. Cambridge: Cambridge University Press.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.

Robinson, P. (1995). Aptitude, awareness and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 303–358). Manoa, HI: University of Hawai'i Press.

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2013). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247-267). Abingdon: Routledge.

Roediger, H., Zaromb, F., & Lin, W. (2017). A typology of memory terms. In B. J. H (Ed.), *Learning and memory: A comprehensive reference* (Vol. 1, pp. 7-19). Amsterdam: Academic press.

Rogers, J., Revesz, A., & Rebuschat, P. (2015). Challenges in implicit learning research: Validating a novel artificial language. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 275-301). Amsterdam, The Netherlands: John Benjamins.

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does Modality Matter? The Effects of Reading, Listening, and Dual Modality on Comprehension. *SAGE Open, 6*(3), 2158244016669550. doi:10.1177/2158244016669550

Rollins, H. A., & Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. *Journal of Experimental Psychology-Human Perception and Performance, 6*(1), 99-109. doi:10.1037/0096-1523.6.1.99

Rost, M. (1990). *Listening in language learning*. London: Longman.

Rummer, R., & Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language, 45*(3), 451-467.

Rummer, R., & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly Journal of Experimental Psychology Section A, 56*(1), 83-95.

Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge university press.

Sagarra, N., & Abbuhl, R. (2013). Optimizing the noticing of recasts via computer-delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *The Modern Language Journal, 97*(1), 196-216.

Sakai, H. (2018). Working memory in listening. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-6): John Wiley & Sons, Inc.

Sanz, C. (2005). *Mind and context in adult second language acquisition methods, theory, and practice*: Georgetown University Press.

Sanz, C., Lin, H.-J., Lado, B., Stafford, C. A., & Bowden, H. W. (2014). One size fits all? Learning conditions and working memory capacity in Ab Initio language development. *Applied Linguistics, 37*(5), 669-692. doi:10.1093/applin/amu058

Sarandi, H. (2015). Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond…? *Language Testing, 32*(4), 485-501.

Schachter, J. (1989). Testing a proposed universal. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 73-88). Cambridge: Cambridge University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1992). Awareness and second language acquisition. *Annual review of Applied Linguistics, 13*(1), 206-226.

Schmidt, R. (2001). Attention. In *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schulz, E., Krahl, G., & Reuschel, W. (2000). *Standard Arabic: An elementary-intermediate course*. Cambridge: Cambridge University Press.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163.

Segalowitz, N. (2003). Automaticity and second languages. In M. H. Long & C. J. Doughty (Eds.), *The handbook of second language acquisition* (pp. 382-408). Oxford: Blackwell.

Segalowitz, N. (2013). Automaticity. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 53-57). London: Routledge.

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks. In A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 157-167). New York: Academic Press.

Shallice, T., & Vallar, G. (1990). The impairment of auditory-verbal short-term storage. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 11-53). Cambridge: Cambridge University Press.

Sharwood-Smith, M. (1981). Consciousness-raising and the second language learner. *Applied linguistics, 2*(2), 159-168.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190. doi:10.1037/0033-295X.84.2.127

Shin, J., Dronjic, V., & Park, B. (2019). The interplay between working memory and background knowledge in L2 reading comprehension. *TESOL Quarterly, 53*(2), 320-347. doi:10.1002/tesq.482

Song, M.-J., & Suh, B.-R. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System, 36*(2), 295-312. doi:10.1016/j.system.2007.09.006

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: comparisons with other validation studies. *Language Learning, 65*(3), 723-751.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308. doi:10.1111/j.1467-9922.2010.00562.x

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*. doi:10.1037/h0093759

Sun, C.-F., & Givón, T. (1985). On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language, 61*(2), 329-351. doi:10.2307/414148

Suzuki, Y., & DeKeyser, R. (2015). Comparing Elicited Imitation and Word Monitoring as Measures of Implicit Knowledge. *Language Learning, 65*(4), 860-895. doi:10.1111/lang.12138

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), (pp. 35-53). Cambridge, MA: Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinke (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–481). Mahwah, NJ: Lawrence Erlbaum.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A steptowards second language learning. *Applied Linguistics, 16*(3), 371–391.

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems* (Vol. 1). Cambridge: Cambridge University Press.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50-73.

Tagarelli, K. M., Mota, M. B., & Rebuschat, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In Z. E. Wen, M. B. Mota, & A. McNeill

(Eds.), *working memory in second language acquisition and processing* (pp. 224-247). Bristol: Multilingual Matters.

Thompson, S. A., & Li, C. N. (2009). The world's major languages. In B. Comrie (Ed.), *The world's major languages* (pp. 703-723). Abingdon: Routledge.

Thuy-Minh, N., Hanh, P., & Pham, T. (2012). The relative effects of explicit and implicit form-focused instruction on the development of L2 pragmatic competence. *Journal of Pragmatics, 44*(4), 416-434. doi:10.1016/j.pragma.2012.01.003

Tomlin, R. (2014). *Basic word order: Functional principles*.

Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language learning, 56*(2), 319-385.

Treiman, R. (2003). Linguistics and reading. In M. Aronoff & J. Rees-Miller (Eds.), *The handbook of linguistics* (pp. 664-672). Oxford: Blackwell.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological review, 76*(3), 282-299.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154. doi:https://doi.org/10.1016/0749-596X(89)90040-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498-505. doi:10.3758/BF03192720

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching, 40*(3), 191-210.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning, 65*(2), 390-416. doi:10.1111/lang.12105

Vandergrift, L., & Goh, C. (2012). Teaching and learning second language listening: Metacognition in action. *Teaching and Learning Second Language Listening: Metacognition in Action*, 1-316. doi:10.4324/9780203843376

VanPatten, B. (1990). Attending to form and content in the input. *Studies in second language acquisition, 12*(03), 287-301.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

VanPatten, B. (2004). Input and output in establishing form-meaning connections. In B. VanPatten, J. N. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 29-47). Mahwah: Lawrence Erlbaum.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*(2), 225–243.

VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In *Second Language Acquisition Theory and Pedagogy* (pp. 169-185). Mahwah, NJ: Lawrence Erlbaum.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics, 12*(1), 54-73.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics, 25*(3), 315-339. doi:10.1093/applin/25.3.315

Wang, M. (1988). Comments on Sun and Givon's study of the OV constructions in Mandarin. *Journal of the Chinese Language Teacher's Association*, 23-33.

Wang, Y. (2017). *English transitivity alternation in second language acquisition: An attentional approach*. London: Routledge.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics, 22*(1), 1-22.

Wen, Z. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. E. Wen & M. B. Mota (Eds.), *Working memory in second language acquisition and processing* (pp. 41-62). Bristol: Multilingual Matters.

White, L. (1988). Island effects in second language acquisition. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition* (pp. 144-172). Dordrecht: Springer.

Williams, J., & Doughty, C. J. (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-353). Bingley: Emerald.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua, 118*(4), 522-553. doi:10.1016/j.lingua.2007.03.003

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition, 23*(03), 345-368.

Wu, Y. (1998). What do tests of listening comprehension test? - A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing, 15*(1), 21-44. doi:10.1177/026553229801500102

Yule, G. (1998). *Explaining English Grammar: A Guide to Explaining Grammar for Teachers of English as a Second Or Foreign Language*. Oxford: Oxford University Press.

Zhang, R. H. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37*(3), 457-486. doi:10.1017/s0272263114000370

References of photos

Elicited imitation task (Version A)

1-1 Practice
Retrieved from: http://clipart-library.com/clipart/n3572876.htm

1-2 Practice
Retrieved from: https://fullskillsexamprep.com/blog/2016-iccrc-full-skills-exam-dates-for-rcics-and-risias/

2-1 pracrice
Retrieved from: https://www.jsib.co.za/3-tips-to-prevent-an-insurance-claim/

2-2 practice
Retrieved from: https://www.hotpoint.co.uk/h/laundry/washing-machines/hotpoint-smart-wmfug-942g-washing-machine-graphite/f086254

3-1 practice
Retrieved from: https://www.american-rails.com/sdp40f.html

3-2 practice
Retrieved from: https://www.thenewstribe.com.au/2015/05/13/worlds-largest-limousine-has-helipad-on-it/

1-1
Retrieved from: https://www.india.com/travel/articles/3-years-around-the-world-the-best-travel-video-that-you-will-ever-watch-3238538/

1-2
Retrieved from: https://www.dreamstime.com/royalty-free-stock-image-tourism-concept-eiffel-tower-over-earth-globe-white-background-image32638256

2-1
Unavailable

2-2
Retrieved from: https://blog.hubspot.com/marketing/bad-habits-for-productivity

3-1
Retrieved from: https://www.theguardian.com/money/blog/2011/nov/11/second-living-room-garage

3-2
Retrieved from: https://logcabins.co.uk/keops-interlock-garages-workshops/

4-1
Retrieved from: https://www.netdoctor.co.uk/parenting/baby-and-toddler/a27185/how-to-wean-baby/

4-2
Retrieved from: https://www.weekendnotes.com/chick-flicks-to-make-you-cry/

5-1
Retrieved from: https://www.cbc.ca/news/canada/calgary/math-teacher-defends-alberta-curriculum-1.2501009

5-2
Retrieved from: http://www.britishielts.in/blog/how-to-choose-best-ielts-course-for-beginners/

6-1
Retrieved from: http://www.extraordinaryeducationbyaurora.com/

6-2
Retrieved from: http://glenbrienns.scoilnet.ie/blog/2013/11/

7-1
Retrieved from: https://www.gettyimages.ae/detail/photo/happy-man-in-office-royalty-free-image/540796331

7-2
Retrieved from: https://www.eduinreview.com/blog/2012/09/employment-rates-and-average-salaries-down-for-new-college-grads/

8-1
Retrieved from: http://www.amlmt.com.br/empresa.php

8-2
Unavailable

9-1
Retrieved from:
https://www.google.com/imgres?imgurl=https://sfpl.org/images/about/blip/presidio/presidio_branch_opening-231.jpg&imgrefurl=http://sfpl.org/?pg=2000348001&h=307&w=460&tbnid=ru4vajWoaNnd9M&tbnh=183&tbnw=275&usg=K_QM_oLeGCrzb_a64L0pGt-fznDvU=&hl=en-SA&docid=XiP9epZVLfAZ_M

9-2
Retrieved from: https://www.gettyimages.ae/videos/reading-candles?sort=mostpopular&phrase=reading%20candles

10-1
Retrieved from: http://www.jejassociates.com/2013/03/

10-2
http://neatoday.org/2015/02/23/school-cell-phone-bans-end-educators-weigh-pros-cons/

11-1
Retrieved from: https://prettyvacant.co/category/university-life/

11-2
Retrieved from: https://www.news.com.au/travel/travel-updates/airport-bars-doing-roaring-trade-from-thirsty-aussie-travellers/news-story/815a76f388202dedfd9ead5586aec661

12-1
Retrieved from: https://www.trendsnhealth.com/weird-hairstyles/6/

12-2
Retrieved from: https://www.nicehair.org/hair-transplants-go-wrong/

13-1
Retrieved from: https://www.google.com/imgres?imgurl=https://images-na.ssl-images-amazon.com/images/I/31xJsNk4TLL.jpg&imgrefurl=https://www.amazon.com/Search-Rescue-Dog-Collar-Enforcement/dp/B00YX58BEK&h=276&w=200&tbnid=zz4_FE-l6GpW5M&tbnh=264&tbnw=191&usg=K_SebBhBmBvywdigbiBYdf156QQkI=&hl=en-SA&docid=zu77QArnmWKnKM

13-2
Retrieved from: https://www.theguardian.com/books/2014/jul/29/cat-sense-john-bradshaw-review

14-1
Retrieved from: https://www.thesun.co.uk/news/1265310/need-to-see-a-gp-the-average-waiting-time-for-an-appointment-is-now-13-days/

14-2
Retrieved from: http://johngreathouse.com/bank-robbery/

15-1
Retrieved from: https://www.alamy.com/stock-photo/messy-conference-room.html

15-2
Unavailable

16-1
Retrieved from: https://www.indiamart.com/mothers-intl/services.html#maths-classes

16-2
Retrieved from: https://education.seattlepi.com/advantages-disadvantages-different-learning-styles-3581.html

17-1
Retrieved from: https://www.tes.com/lessons/EUbdy-CljqGeFg/sound-of-rain

17-2
Unavailable

18-1
Retrieved from: https://www.arts.ac.uk/colleges/london-college-of-communication/stories/lcc-publishing-students-present-to-ipc-media-panel

18-2
Retrieved from: https://news.stlpublicradio.org/post/school-transfer-law-should-include-private-school-option-lawmakers-say#stream/0

19-1
Retrieved from: https://www.aastweb.org/blog/10-proven-tips-to-get-your-sleep-deprived-teen-sleep-better

19-2
Retrieved from: http://192.168.8.1/html/home.html

20-1
Retrieved from: https://abcnews.go.com/US/spider-caused-car-crash/story?id=33926515

20-2
Retrieved from: https://www.zastavki.com/eng/Auto/Kia/wallpaper-71406.htm

21-1
Unavailable

21-2
Retrieved from: https://www.chronicbodypain.net/treat-your-fibromyalgia-naturally/

22-1
Retrieved from: https://cygnetnewtheatre.com/exterior-home-color-visualizer/inspiring-exterior-home-color-visualizer-3/

22-2
Retrieved from: https://www.servicemaster-emt.com/las-vegas-nv/disaster-restoration/

23-1
Retrieved from: https://overabundantcuriosity.wordpress.com/2012/05/27/last-years-garden-project/

23-2
Unavailable

24-1
Retrieved from: https://uk.news.yahoo.com/man-eaten-sausages-chips-22-074936960.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACyzOHu-w5eBihpmkrK2Ptpmlzd8DPYN8KFwjPPIy4u8iZ8VuHDd1yjxcn5t8TKLqFrhoyTFuoinvjLDP9RrMpZWSEw1LdcMuNgFaNXKAGBGwh0HorK-DKc47VeONqykSd7p2Wq5Lod8B2rLhABlxrB6UjFrUshDiYVsbmu1kvqo

24-2
Retrieved from: http://friartucksfleetingthoughts.blogspot.com/2013/09/hospital-for-sinners-sermon-labor-day.html

25-1
Retrieved from: https://work.chron.com/personality-traits-become-secretary-30993.html

25-2
Retrieved from: http://nightcelestine.blogspot.com/2010/07/chapter-iv-what-good-teachers-are-made_26.html

26-1
Retrieved from: https://www.shutterstock.com/video/clip-1845961-close-up-attractiv-young-photographer-taking-photographs--

26-2
Unavailable

27-1
Retrieved from: http://slworkshop.net/2015/04/changes/

27-2
Retrieved from: https://www.hmfh.com/projects/east-fairhaven-elementary-school/

28-1
Unavailable

28-2
Retrieved from: https://www.history.org/Foundation/journal/Christmas05/Food.cfm

29-1
Retrieved from: https://www.flickr.com/photos/galapagos-spanish-school/3293129915/

29-2
Retrieved from: https://www.gmridiomas.com/es/cursos/academia/reino-unido/liverpool/school-of-english

30-1
Retrieved from: http://www.letribunaldunet.fr/photos/20-signes-montrent-etes-vieux-corps-jeune.html/3?wptouch_shortcode=1&current_time=1561766400095

30-2
Retrieved from: https://mylife-is-mylove.blogspot.com/2013/09/mamas-love.html

31-1
Retrieved from: https://chicago.suntimes.com/2016/7/5/18325587/people-unplug-for-table-tennis-in-millennium-park
31-2

Retrieved from: https://guci.org/2013/07/22/off-the-rim/

32-1
Retrieved from: https://www.cameratim.com/reviews/computing/asus-pro31j-series-f3jc-laptop/

32-2
Retrieved from: https://sp.depositphotos.com/92345230/stock-photo-happy-relaxed-young-woman-chatting.html

Focus on form task

1-1
Retrieved from: http://www.ohwhatabeautifulgarden-chicagoland.com/?p=8647

1-2
Retrieved from: https://amazing-pics-hub.com/lovely-valentines-day-wallpapers-for-2016/

2-1
Retrieved from: https://bothsidesofthetable.com/never-ask-a-busy-person-to-lunch-here-s-why-94fa63580354?gi=28595181bfdb

2-2
Retrieved from: https://meghan03996.wordpress.com/2015/09/23/visual-stereotypes-how-stereotypes-are-used-today-blog-post-5/

3-1
Retrieved from: https://wordsareavenues.wordpress.com/2012/07/30/lacking-profundity/

3-2
Retrieved from: https://www.telegraph.co.uk/motoring/news/8720475/Bank-holiday-getaway-could-be-smoother-than-normal.html

4-1
Retrieved from: http://www.myaspergerschild.com/2011/08/dealing-with-aspergers-children-who.html

4-2
Retrieved from: http://lettoysbetoys.org.uk/page/4/?s=reading

5-1
Retrieved from: http://chegueiaomundo.com.br/bebes/cardapios-saudaveis-para-festas-infantis/

5-2
Retrieved from: https://atlantablackstar.com/2015/08/30/university-pennsylvania-study-says-school-suspensions-black-children-south/

6-1

Retrieved from: https://www.bgdblog.org/2013/12/calling-less-disposable-way-holding-accountable/

6-2
https://www.considerable.com/life/communication/dealing-dysfunctional-family/

7-1
Unavailable.

7-2
Retrieved from: http://www.childwisechat.com/2012/07/27/see-my-rain-gutter-bookshelves/

8-1
http://elderaffairs.state.fl.us/doea/elder_helpline.php

8-2
Retrieved from: https://www.onsip.com/voip-resources/voip-reviews/voip-comparison-top-13-questions-you-should-ask-list

9-1
Retrieved from: https://www.123rf.com/photo_18635821_smiling-waiter-taking-order-from-men-customers-outdoor-bar.html

9-2
Retrieved from: http://kindergartensmarts.com/2015/08/i-have-that-kid-in-my-classroom.html

10-1
Retrieved from: http://superjellyhoneywall.blogspot.com/2012/07/easy-jams.html

10-2
Retrieved from: https://casaveneracion.com/cheesecake-with-homemade-blueberry-topping/

11-1
Retrieved from: http://strakul.blogspot.com/2012_01_01_archive.html

11-2
Retrieved from: https://wall.alphacoders.com/big.php?i=304012

12-1
Retrieved from: https://www.lawinprocess.com/interview-questions-may-want-avoid/

12-2
Retrieved from: https://www.gettyimages.ae/detail/news-photo/senator-jim-demint-talks-on-the-phone-in-his-office-news-photo/157779977

13-1
Retrieved from: https://splus.resmed.com/feeling-better-less-deep-sleep/

13-2

Retrieved from: https://www.theepochtimes.com/causes-of-groin-pain-and-swelling-in-men_1917710.html

14-1
Retrieved from: http://www.shutterstock.com/pic-135640697/stock-vector-police-man-in-the-black-uniform-caught-the-criminal-police-officer-arrested-thief.html

14-2
Retrieved from: https://www.runnersworld.com/runners-stories/a20802921/runners-run-down-would-be-jewel-thief/

15-1
Retrieved from: https://www.huffpost.com/entry/brain-development-poverty_n_6968758

15-2
Retrieved from: https://mybindi.typepad.com/parenting/2013/11/how-to-start-giving-your-child-an-allowance.html

16-1
Retrieved from: http://www.daehanmindecline.com/2013/20130612.html

16-2
Retrieved from: https://www.dreamstime.com/stock-photography-raining-heavily-image18928622

17-1
Retrieved from: https://www.liputan6.com/health/read/794962/kenapa-membedong-bayi-terlalu-lama-tak-baik

17-2
Retrieved from: https://www.myboatlife.com/2012/03/bed-options-for-newborn-infant-or-baby.html

18-1
Retrieved from: https://www.dailymail.co.uk/news/article-2290646/Jean-splashes-brand-new-car-age-94-shes-speeding-ticket-80-years.html

18-2
Retrieved from: https://vimeo.com/11736134

19-1
Retrieved from: https://rxcomfort616.com/furnace-repair-grand-rapids-mi/

19-2
Retrieved from: https://www.dailymail.co.uk/femail/article-3425209/Is-world-s-warmest-coat-Stuffed-rare-ox-hair-900-jacket-said-snug-wool-FEMAIL-braves-elements-test.html

20-1
Retrieved from: https://97zokonline.com/6-things-you-can-do-so-you-dont-get-sick/

20-2
Retrieved from: https://giddyinthecityblog.wordpress.com/2014/10/18/anatomy-of-a-hip-dinner-party/

21-1
Retrieved from: https://pt.depositphotos.com/1833543/stock-photo-two-sad-children-on-steps.html

21-2
Retrieved from: https://loyalty360.org/content-gallery/daily-news/papa-johns-fans-can-earn-loyalty-points-by-watching-tv

22-1
Retrieved from: https://www.ubergizmo.com/2011/06/spiderpro-holster/

22-2
Retrieved from: https://photoh.com.au/full-frame-versus-sensor-crop-cameras-whats-best-for-you/

23-1
Retrieved from: https://totalfood.com/cass-calder-smith-of-ccs-architecture-new-york-ny-and-san-francisco-ca/

23-2
Retrieved from: https://www.yelp.co.uk/search?cflt=cafes&find_near=st-pauls-square-birmingham

24-1
Retrieved from: https://www.istockphoto.com/vector/confused-child-shrugging-shoulders-gm477249514-66756599

24-2
Retrieved from: http://www.wanee.org/ParentsandStudents.aspx

25-1
Retrieved from: https://www.5minutesformom.com/48716/exercising-as-a-whole-family-guest-post/

25-2
Retrieved from: https://www.chrishecker.com/Achievements_Considered_Harmful%3F

26-1
Retrieved from: https://www.zillow.com/levittown-ny/colonial-house_att/

26-2
Retrieved from: https://www.zillow.com/levittown-ny/colonial-house_att/

27-1

Retrieved from: http://www.catoverload.com/sehr-haesslicher-tiger-ugly-tiger/

27-2
Retrieved from: http://animalhdp.blogspot.com/2015/01/the-bengal-tiger-is-national-animal-of.html

28-1
Retrieved from: https://texashillcountry.com/man-carrying-an-american-flag-across-west-texas/

28-2
Retrieved from: https://www.dailymail.co.uk/sciencetech/article-2696136/Why-people-Denmark-happy-Study-claims-Danish-DNA-king-pursuit-happiness.html

29-1
Retrieved from: https://www.pinterest.com/baumwoolgerhome/

29-2
Unavailable

30-1
Unavailable

30-2
Retrieved from: http://thebiz.bentley.edu/11-essential-apps-for-college-students/

31-1
Unavailable

31-2
Retrieved from: https://www.express.co.uk/news/nature/499381/UK-weather-Final-storms-rain-and-gales-from-Hurricane-Bertha

32-1
Retrieved from: https://phz8.petinsurance.com/healthzone/pet-health/pet-toxins/garlic-toxicity-and-pets

32-2
Retrieved from: https://commons.wikimedia.org/wiki/File:Orange-Whole-%26-Split.jpg

33-1
Retrieved from: https://www.huffpost.com/entry/a-new-way-to-pay-online-c_b_168340

33-2
Retrieved from: https://davelearningstuff.wordpress.com/2014/07/01/jrr-tolkien-typed-the-lord-of-the-rings-with-two-fingers/

34-1

Retrieved from: https://depositphotos.com/30396331/stock-photo-man-learning-english-isolated-on.html

34-2
Retrieved from: https://www.brilliantpublications.co.uk/book/verbarama-futur-proche-722

35-1
Retrieved from: http://www.home-designing.com/2010/08/modern-style-kitchen-designs

35-2
Retrieved from: http://http//www.cheverellwood.co.uk/kitchens/traditional/

36-1
Retrieved from: https://www.wisegeek.com/what-is-a-polar-bear-plunge.htm

36-2
Retrieved from: http://www.camelbackcoaching.com/blog-news/20-calming-techniques-for-open-water-swimming/

Elicited imitation task (Version B)

1-1
Retrieved from: https://dadischeap.com/2015/10/08/the-perks-of-driving-an-older-car/

1-2

Retrieved from: https://www.autosmodel.com/what-are-you-looking-for-in-a-brand-new-car/brand-new-car/

2-1
Retrieved from: https://www.bgs.org.uk/blog/a-bespoke-blue-light-response-to-frail-older-fallers-makes-complete-sense-%E2%80%93-but-does-it-work

2-2
Unavailable

3-1
Retrieved from: https://www.gettyimages.ae/detail/photo/cute-daughter-and-her-father-sleeping-in-one-bed-royalty-free-image/467745960

3-2
Retrieved from: http://www.herbawi.com/en/1/7/43

4-1
Retrieved from:
https://www.google.com/imgres?imgurl=https://www.upweydoctors.com.au/wp-content/uploads/2019/09/image-2.png&imgrefurl=https://www.upweydoctors.com.au/uncategorized/814/&h=1024&w=683&

tbnid=HMtxvcz5-ohaBM&tbnh=275&tbnw=183&usg=K_QpvuEm-t38d0OUHxOfX__JlcPlw=&hl=en-SA&docid=PEfyg7Q_s_HICM

4-2
Retrieved from:
https://www.google.com/imgres?imgurl=https://farm4.static.flickr.com/3894/15216831239_0211f1afcf.jpg&imgrefurl=https://www.flickriver.com/photos/dianachaloux/tags/hitchfit/&h=468&w=499&tbnid=l8XWnZR8zbcZkM&tbnh=217&tbnw=232&usg=K_mm04PVmyngNztpDRPLG5sR4d81w=&hl=en-SA&docid=R8OfiMoLowWwRM

5-1
Retrieved from:
https://commons.wikimedia.org/wiki/File:Ellen_H._Swallow_Richards_House_Boston_MA_01.jpg

5-2
Retrieved from: https://labs.sogeti.com/lets-help-around-and-save-the-society-with-a-cool-app/

6-1
Retrieved from: https://astridlindgrenmemorialaward.wordpress.com/2013/06/12/in-astrid-lindgrens-living-room/

6-2
Retrieved from: https://blogs.gwu.edu/postcardsfromgw/2014/02/05/happy-chinese-new-year-and-superbowl/

7-1
Retrieved from: https://www.theguardian.com/film/2013/dec/02/cinema-paradiso-25th-anniversary-salvatore-cascio

7-2
Retrieved from: https://www.vaterfreuden.de/tipps/freizeittipp/mit-kind-ins-kino8-1

8-1
Retrieved from: https://www.wired.com/2010/01/apple-january-event/

8-2
Retrieved from: https://www.engadget.com/2007/05/30/palm-foleo-announced/

9-1
Retrieved from: https://www.news.com.au/finance/supermarket-milk-war-turns-sour-after-woolworths-pledges-cash-to-dairy-farmers/news-story/2f926463f844466f2ddcbab3dbe1ca8e?sv=a83345035ae3cc927bb8f2effbe60667

9-2
Retrieved from: http://www.eoinbutler.com/home/what-peaches-and-what-penumbras-whole-families-shopping-at-night-2/

10-1
Retrieved from:
https://www.google.com/imgres?imgurl=https://thumbs.dreamstime.com/t/man-big-bread-13388425.jpg&imgrefurl=https://www.dreamstime.com/photos-images/big-bread.html&h=160&w=107&tbnid=FSI5FnEIlIXIuM&tbnh=160&tbnw=107&usg=K_0Vc6FcF751TpyrsjF8X6Q-BkX1U=&hl=en-SA&docid=4k7vxc6x_3Q9TM

10-2
Retrieved from: http://s697.photobucket.com/user/Biorihime/media/erza/erza-armor2.gif.html?o=21

11-1
Retrieved from:
https://www.google.com/imgres?imgurl=https://img.medscape.com/thumbnail_library/dt_150319_angry_argue_doctors_800x600.jpg&imgrefurl=https://www.medscape.com/viewarticle/845790_7&h=600&w=800&tbnid=hJwRmulH_E-BDM&tbnh=194&tbnw=259&usg=K_UTEOLcfC48eSe2S8hVxduMEPLZI=&hl=en-SA&docid=FJae-Sgs859Q3M

11-2
Retrieved from: https://www.tellwut.com/surveys/lifestyle/living/85556-chronic-complainers.html

12-1
Retrieved from: http://mostfunnyimags.blogspot.com/2011/09/funny-girl-funny-face.html

12-2
Retrieved from:
https://www.google.com/imgres?imgurl=https://thumbs.dreamstime.com/b/funny-man-watermelon-helmet-googles-looks-like-parasitic-caterpillar-34231117.jpg&imgrefurl=https://www.dreamstime.com/stock-photography-bizarre-man-helmet-watermelon-riding-fast-image34231122&h=677&w=800&tbnid=W9DAD9ERZHovtM&tbnh=206&tbnw=244&usg=K_BtF2IsSzSjadqqCrFLO5ovQthZ0=&hl=en-SA&docid=6132TxVtMUM4PM

13-1
Retrieved from:
https://www.google.com/imgres?imgurl=https://imagesvc.meredithcorp.io/v3/mm/image%3Furl%3Dhttps://static.onecms.io/wp-content/uploads/sites/38/2015/06/12232158/szZXFoNDE68E36WiX9z7ka6Q2twv02AZ.png&imgrefurl=https://www.parents.com/kids/discipline/setting-limits/getting-kids-to-follow-the-rules/&h=506&w=900&tbnid=2MvQUBpH20bImM&tbnh=168&tbnw=300&usg=K_Tg9PWG6Yr6c8ZIrOOjA5n5Jr0GQ=&hl=en-SA&docid=XrFUcp8ODdWZGM

13-2
Retrieved from: https://thepeakcounselinggroup.org/understanding-the-complex-relationship-between-controlling-parents-and-who-children-that-love-them/

14-1
Retrieved from: https://www.theguardian.com/lifeandstyle/2016/jun/19/reality-tv-is-dead-ruined-by-reality-itself-eva-wiseman

14-2
Unavaialable

15-1
Retrieved from: http://thincdesign.com/project/discovery-center

15-2
Retrieved from: https://hnc3d.wordpress.com/2015/03/30/hello-world/

16-1
Retrieved from: https://www.cnbc.com/2015/08/05/anger-in-the-office-costs-women-more-than-men.html

16-2
Retrieved from: http://www.westshoreprimarycare.com/blog/category/Endocrinology

17-1
Retrieved from: https://zarembalaw.com/happy-birthday-so-how-well-are-you-driving-lately/

17-2
Retrieved from: http://www.idioms4you.com/complete-idioms/no-skin-off-my-nose.html

18-1
Unavailable

18-2
Retrieved from: https://thenextweb.com/insider/2013/04/11/andreessen-horowitz-brings-on-tech-veteran-ken-coleman-as-special-advisor-to-turn-founders-into-ceos/

19-1
Unavailable

19-2
Retrieved from: https://www.thecourier.com.au/story/3038250/driver-crashes-into-fence/

20-1
Retrieved from: https://pantip.com/topic/34658704

20-2
Retrieved from: https://emcomail.com/templates/business-card/envelope-template/

21-1
Retrieved from: http://garden-designer.org/2016/09/

21-2
Retrieved from: file:///C:/Users/Sami/Desktop/Photo_references/Post-test%20Change/21-2.jpg

22-1
Retrieved from: https://www.goodtherapy.org/blog/rainbow-racism-it-doesnt-stop-with-being-gay-1017134

22-2
Retrieved from: https://www.nairaland.com/4091292/facts-unilag-boys

23-1
Retrieved from: https://www.beliefnet.com/faiths/galleries/how-does-faith-work-in-modern-life.aspx

23-2
Unavailable

24-1
Unavailable

24-2
Retrieved from: https://parentingwithunderstanding.com/tag/crying/

25-1
Retrieved from: https://www.usnews.com/education/online-education/articles/2014/10/22/look-at-health-care-trends-growth-to-research-online-nursing-programs

25-2
Retrieved from: http://locatible.com/blog/medical/strangled-with-my-stethoscope-twice-nurses-do-not-feel-safe-at-work/

26-1
Retrieved from: https://biblioteca.regielive.ro/proiecte/geografie/singapore-109092.html

26-2
Retrieved from:
https://www.google.com/imgres?imgurl=https://i.pinimg.com/originals/fd/75/a9/fd75a9156357a9f9dfc723ada5a854e2.jpg&imgrefurl=https://www.pinterest.com/pin/324048135678208241/&h=681&w=1024&tbnid=BSu-In3FFXDSvM&tbnh=183&tbnw=275&usg=K_uf7rDrOQZWwFKkxJj9tNt5yY1Ik=&hl=en-SA&docid=wj6CSX0WH_UjlM

27-1
Retrieved from: https://www.google.com/imgres?imgurl=http://3.bp.blogspot.com/-OwIaNYlx2CM/VG56DuGK14I/AAAAAAAAAVs/-KOyIVzndJo/s1600/HIGH%252BPRESSURE.jpg&imgrefurl=http://lavadodecoches.blogspo

t.com/2014/11/&h=504&w=800&tbnid=f6qIBGNXVWfAnM&tbnh=178&tbnw=283&usg=
K_Y3urw_f5x9cfBVRQ_cIKkapiJpA=&hl=en-SA&docid=AcZdoDkxGeB9tM

27-2
Unavailable

28-1
Retrieved from: https://footage.framepool.com/en/bin/2663342,student,lecturer,conversing/

28-2
Unavailable

29-1
Unavailable

29-2
Retrieved from:
https://www.telegraph.co.uk/finance/newsbysector/constructionandproperty/9098319/Bumpe
r-profits-from-UK-house-builders.html

30-1
Retrieved from:
https://www.google.com/imgres?imgurl=http://media.wiley.com/Lux/13/88513.image4.jpg%
3Fh%3D400%26w%3D535&imgrefurl=https://prezi.com/8is0i1mkkkni/module-6-
summative-
assessment/&h=400&w=535&tbnid=b4ukgl0PtqMhpM&tbnh=194&tbnw=260&usg=K_Thc
spIsyn4sKc8G4wYce0NSnbXg=&hl=en-SA&docid=fxmZlI5dfqoXeM

30-2
Unavailable

31-1
Retrieved from: https://onedio.com/haber/rihanna-afrikali-kizlara-egitim-ve-bisiklet-
saglamak-icin-ofo-yla-yeni-bir-proje-baslatti-780322

31-2
Unavailable

32-1
Retrieved from: https://egpsac.com/blog/a-maintenance-of-a-business-continuity-
management-system-a-managerial-approach/

32-2
Retrieved from: http://innergise.com.au/2016/08/4ir-challenges-engineers/

306

Appendix

1. Stimuli in training tasks

1.1. Focus on form task

	Structure	Stimulus
1	NA	Seldom have I seen such beautiful roses.
2	NA	Rarely does John refuse my invitation to dinner.
3	NA	Seldom have the boys experienced such rain.
4	NA	Rarely can the child accept that it is time to leave.
5	NA	Rarely do the students have a birthday party.
6	NA	Rarely did my grandparents argue with each other.
7	NA	Rarely did the children clean their rooms.
8	NA	Seldom will my parents phone before visiting me.
9	NA	Rarely does the waiter smile at the customers.
10	NA	Seldom does Sally make homemade cheesecake.
11	NA	Rarely can I remember seeing such an amazing sunset.
12	NA	Seldom has David spoken with me softly at his office.
13	CON	Had you gone to bed early, you would have felt great.
14	CON	Had the police arrived later, the thief would have run away.
15	CON	Had you completed the task, I would have paid you.
16	CON	Had the rain stopped, we would have gone out for a walk.
17	CON	Had the child woken in the night, we would have calmed him.
18	CON	Had Helen bought the car, she would have come to see us.
19	CON	Had Nina bought a coat, she would have felt warmer.
20	CON	Had I felt better, I would have come to your party.
21	CON	Had the children left, they would have missed the show.
22	CON	Had the camera worked well, I would have taken it.
23	CON	Had the café looked cheap, we would have gone in.
24	CON	Had the boy found the bus stop, he would have arrived on time.
25	TQ	The children have been playing nicely, haven't they?
26	TQ	House prices in the US are not very high, are they?
27	TQ	The tiger in the small zoo is not attractive, is it?
28	TQ	You have travelled to Denmark several times, haven't you?

29	TQ	The company has not finished building the house, has it?
30	TQ	The students are working hard now, aren't they?
31	TQ	The weather does not look great today, does it?
32	TQ	Your child is allergic to garlic, isn't he?
33	TQ	Martin has typed the article with two fingers, has he?
34	TQ	Alex is in Canada to learn French, isn't he?
35	TQ	Alan and his wife prefer the traditional kitchen, don't they?
36	TQ	Jack is the first British man to swim in the icy water, isn't he?

The following are screenshots of one trial for the focus on form task:

Slide 1:

Welcome to the
Computer Lab

Start the task by pressing the space bar
when you are ready.

Slide 2:

In this task, you will hear a number of sentences presented individually.
Each sentence will be followed with a page showing two pictures in that
one of them describes the sentence you have just heard.

You must decide which picture best describes the sentence by pressing
either the 'left' or 'right' keys on the keyboard. If you are not sure about
your answer, please press the 'down' key as shown in the picture to the
right. After this, you are required to repeat the sentence orally.

Remember that you should press the space bar when you finish saying
the sentence.

Please press the space bar to start the task immediately.
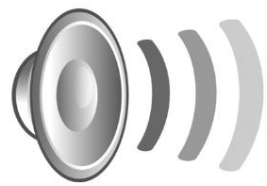
Left   Right
Down
(Not sure)

Slide 3:

<p style="text-align:center;">Sentence 1</p>

Slide 4: (here the participant is hearing the stimulus)

Slide 5: (here the participant is required to choose which picture best describes the sentence he/she has heard.

Slide 6: (here the participant is required to repeat the sentence)

Slide 7:

New sentence

Please press the spacebar when you are ready.

1.1.1. Photos used in the focus on form task:

1-1

1-2

2-1

2-2

3-1

3-2

4-1

4-2

5-1

5-2

6-1

6-2

7-1

7-2

8-1

8-2

9-1

9-2

10-1

10-2

dd

11-1

11-2

12-1

12-2

13-1

13-2

14-1

14-2

15-1

15-2

16-1

16-2

17-1

17-2

18-1

18-2

19-1

19-2

20-1

20-2

21-1

21-2

22-1

22-2

23-1

23-2

24-1

24-2

25-1

25-2

26-1

26-2

27-1

27-2

28-1

28-2

29-1

29-2

30-1

30-2

31-1

31-2

32-1

32-2

33-1

33-2
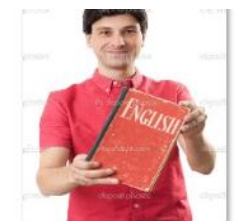
34-1
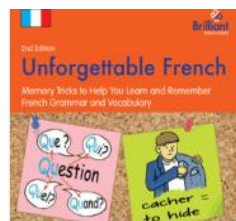
34-2

35-1

35-2

36-1

36-2

1.2. Deductive learning task

	Target structure	Sentence
1	NA	Rarely has Helen finished doing her test early.
2	NA	Seldom will you have the chance to read the novel in the office.
3	NA	Seldom has my friend seen such a music show.
4	NA	Rarely would my wife prepare lunch early on weekdays.
5	NA	Rarely could the children follow their parents' instructions.
6	NA	Seldom have we had such a good time.
7	NA	Seldom did they offer to take care of the baby.
8	NA	Seldom do I keep in touch with my friends through Facebook.
9	NA	Rarely does Jenifer publish two books in a year.
10	NA	Rarely would my husband go to bed before midnight.
11	NA	Seldom have examiners seen such complete answers.
12	NA	Rarely did doctors see such a skin disease.
13	CON	Had I done my best, I would have passed the driving test.
14	CON	Had Tom finished his book on holiday, he would have started another.
15	CON	Had Liza helped me, I would have helped her with her problem.
16	CON	Had my mother known about this, we would have gotten into trouble.
17	CON	Had Nina listened carefully, she would have known the answer.
18	CON	Had the doctor arrived earlier, Sarah would have survived.
19	CON	Had I remembered your birthday, I would have brought you a present.
20	CON	Had we bought flour, we would have made some cakes.
21	CON	Had you received training, you would have done your job better.
22	CON	Had the summer started earlier, we would have enjoyed it more.
23	CON	Had the shop wrapped the present, I would have taken it.
24	CON	Had my friend called earlier, we would have changed our plans.
25	TQ	They are so in love with each other, aren't they?
26	TQ	The tree does not need cutting back completely, does it?
27	TQ	David should bring the cows back home by 5pm, shouldn't he?
28	TQ	You are delivering the baby in week 38, aren't you?

29	TQ	Saudi Arabia and America have similar summer weather, don't they?
30	TQ	The professor died suddenly of a heart attack, didn't he?
31	TQ	Jenifer missed her flight to Japan yesterday, didn't she?
32	TQ	The builder has not come to repair the roof, has he?
33	TQ	Liza can speak German and French, can't she?
34	TQ	This huge building was built 70 years ago, wasn't it?
35	TQ	Helen is not having her hair cut tomorrow, is she?
36	TQ	The boys do not love spicy food sauce, do they?

The following are screenshots of one trial for the deductive learning task:

Slide 1:

This task will consist of several groups of sentences, and each group will target one grammatical rule. The total number of grammatical rules in the whole task are three.

In each of the first three groups, an instruction about one grammatical rule will be provided first, then you will listen to a number of sentences individually targeting this rule.

After that, a question about the sentence will be shown (e.g., Does the sentence you heard contain the word '.......'? You must press the left key for 'No' or the right key for 'Yes' on the keyboard as indicated in the picture to the right. If you are not sure about your answer, please press 'down'.

Press the space bar to continue.

No   Yes   Down (Not sure)

Slide 2:

Finally, you will be required to repeat the sentence orally after the beep.

You must press the space bar when you finish repeating the sentence.

Remember that the instruction about each grammatical rule will be provided only one time in the first three groups, and the other groups will be the same as the first ones, but without instruction.

Do you have any questions?
If not, please press the space bar to start the task.

Slide 3:

Group one

Listen to the sentences carefully and pay attention to where the auxiliary verb comes in each sentence. For example, in the sentence, 'Seldom have they seen such tall buildings', the auxiliary is 'have' and it comes before the subject of the sentence, 'they'.

Please press the space bar to start.

Slide 4:

Sentence 1

Slide 5: (here the participant is hearing the stimulus)

Slide 6:

Does the sentence you heard contain the word 'chance'?

No Yes

Down
(Not sure)
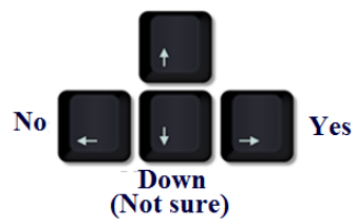
Slide 7: (here the participant is required to repeat the sentence)

Slide 8:

New sentence

Please press the spacebar when you are ready.

2. Stimuli in testing tasks

2.1. Elicited imitation task (version A)

	Structure	Stimulus	Error type
I	TQ	The washing machine is not broken, it is?	Affirmative tag question: S-AUX inversion
II	NA	Seldom had Tom seen such a long train	Correct
III	CON	Had Peter failed the test, he returned home.	Second clause: past form of the V
1	TQ	Lucy has travelled all over the world, has she?	Affirmative-Affirmative
2	TQ	You did not hurry to catch the bus yesterday, didn't you?	Negative-Negative
3	TQ	I should not have a house with two garages, I should?	Affirmative tag question: S-AUX inversion
4	TQ	The little boy needs to be fed now, he doesn't?	Negative tag question: S-AUX inversion

5	TQ	The students are taking a language course, aren't they?	Correct
6	TQ	You remember the chemistry lesson from last week, don't you?	Correct
7	TQ	Peter was happy with the basic salary, wasn't he?	Correct
8	TQ	The babysitters could not meet at the café last week, could they?	Correct
9	NA	Rarely the children do open the book to read at night.	Swapping AUX with S
10	NA	Seldom my classmate would use his phone in class.	Swapping AUX with S
11	NA	Did rarely I travel to the US without my husband.	Swapping NA with AUX
12	NA	Have seldom I seen such a horrible haircut.	Swapping NA with AUX
13	NA	Rarely would the small cat go out at night.	Correct
14	NA	Seldom does a patient wait in the emergency department.	Correct
15	NA	Rarely has the housekeeper cleaned the meeting room.	Correct
16	NA	Seldom will the students enjoy the maths class.	Correct
17	CON	Had the rain stopped, my brother would go to the park.	Second clause: Would + V
18	CON	Had Sally attended the class, she would understand the lesson.	Second clause: Would + V
19	CON	Had David woken up early, he did not miss the bus.	Second clause: Past form of the V
20	CON	Had the car crashed, everyone died.	Second clause: Past form of the V
21	CON	Had my sons visited me, I would have made some cakes.	Correct
22	CON	Had the storm continued, we would have lost the house.	Correct
23	CON	Had Alex cleaned the garden, we would have had a barbecue.	Correct

24	CON	Had Tom called earlier, we would have prepared him a meal.	Correct
25	Filler: Relative clause	This is the teacher which name I always forget.	Swapping 'whose with 'which' (difficult level)
26	Filler: Since/for	I have loved taking photographs from I was a child.	Swapping ' since' with 'from' (difficult level)
27	Filler: present simple	The school library close early at the weekend.	Deleting -s (easy level)
28	Filler: present continuous	Some people still eating in the room.	Deleting 'are' (easy level)
29	Filler: Relative clause	That is the school where I learnt to speak Spanish.	Correct (difficult level)
30	Filler: Since/for	The old man has been sitting here for two hours.	Correct (difficult level)
31	Filler: present simple	My sons play tennis every day after school.	Correct (easy level)
32	Filler: present continuous	Emily is not using the laptop now.	Correct (easy level)

Note. The first three items are practice sentences.

The following are screenshots of one trial for the elicited imitation task:

Slide 1:

In this task, you will hear a number of sentences presented individually.
Each sentence will be followed with a page showing two pictures. One
of them describes the sentence you have just heard.

You must decide which picture best describes the sentence by pressing
either the 'left' or 'right' keys on the keyboard. If you are not sure about
your answer, please press the 'down' key as shown in the picture to the
right. After this, you are required to orally repeat the sentence after the
beep in CORRECT English.

Left   Down (Not sure)   Right

Please press the space bar to continue.

Slide 2:

REMEMBER that you will have a limit time to choose which picture best describes the sentence. Similarly, you will have a limit time to say the sentence in correct English. Therefore, you must do as quickly and accurately as possible.

In the next few minutes, you will do some practice to become familiar with the task.

Please press the space bar to begin the practice trial.

Slide 3:

Sentence 1

Slide 4: (here the participant is hearing the stimulus)

Slide 5: (here the participant is required to choose which picture best describes the sentence he/she has heard.

Slide 6: (here the participant is required to repeat the sentence)

Slide 7:

New sentence

Please press the spacebar when you are ready.

2.1.1. Photos used in the EIT (version A):

1-1

1-2

2-1

2-2

3-1

3-2

4-1

4-2

5-1

5-2

6-1

6-2

7-1

7-2

8-1

8-2

9-1

9-2

10-1

10-2

11-1

11-2

12-1

12-2

13-1

13-2

14-1

14-2

15-1

15-2

16-1

16-2

17-1

17-2

18-1

18-2

19-1

19-2

20-1

20-2

21-1

21-2

22-1

22-2

23-1

23-2

24-1

24-2

25-1

25-2

26-1

26-2

27-1

27-2

28-1

28-2

29-1

29-2

30-1

30-2

31-1

31-2

32-1

32-2

P1 -1practice_sentence

P1 -2practice_sentence

P2 -1practice_sentence

P2 -2practice_sentence

P3 -1practice_sentence

P3 -2practice_sentence

2.2. Elicited imitation task (Version B)

	Structure	Stimulus	Error type
I	TQ	The washing machine is not broken, it is?	Affirmative tag question: S-AUX inversion
II	NA	Seldom had Tom seen such a long train	Correct
III	CON	Had Peter failed the test, he returned home.	Second clause: Past form of the V
1	TQ	Your manager has gotten a brand new car, has he?	Affirmative-Affirmative
2	TQ	You did not call the police last night, didn't you?	Negative-Negative
3	TQ	Your boy should not sleep beside you, he should?	Affirmative tag question: S-AUX inversion
4	TQ	Regular exercise makes you healthier, it doesn't?	Negative tag question: S-AUX inversion
5	TQ	Those neighbours are moving to a new house tomorrow, aren't they?	Correct
6	TQ	The guests need much more space now, don't they?	Correct
7	TQ	The cinema was packed with children, wasn't it?	Correct
8	TQ	You could not lend your iPad to me by any chance, could you?	Correct
9	NA	Rarely the shops do run out of milk in this city.	Swapping AUX with S
10	NA	Seldom you would find this size of bread in the shops.	Swapping AUX with S
11	NA	Did rarely the two teachers argue with each other.	Swapping NA with AUX
12	NA	Have seldom my children met such a funny man.	Swapping NA with AUX
13	NA	Rarely would the boy listen to advice from others.	Correct
14	NA	Seldom do we get the chance to watch television together.	Correct

15	NA	Rarely have the international tourists visited the art gallery.	Correct
16	NA	Seldom will the discussion turn into a big argument.	Correct
17	CON	Had David driven carefully, he would not lose his licence.	Second clause: Would + V
18	CON	Had you come to the meeting yesterday, you would meet Sarah.	Second clause: Would + V
19	CON	Had the car stopped earlier, it did not crash into the fence.	Second clause: Past form of the V
20	CON	Had you given me your home address, I sent you the document.	Second clause: Past form of the V
21	CON	Had the sofa looked new, I would have bought it.	Correct
22	CON	Had my friend come earlier, we would have gone shopping.	Correct
23	CON	Had we gone to the party, we would have taken some desserts.	Correct
24	CON	Had the child eaten his food, he would not have felt hungry.	Correct
25	Filler: Relative clause	I have met the nurse which husband died.	Swapping 'whose with 'which' (difficult level)
26	Filler: Since/for	It has gotten colder from the sun went in.	Swapping ' since' with 'from' (difficult level)
27	Filler: present simple	David usually wash the family car.	Deleting -s (easy level)
28	Filler: present continuous	Some students having a meeting with the tutor now.	Deleting 'are' (easy level)
29	Filler: Relative clause	The UK is a country where many researchers have studied.	Correct (difficult level)
30	Filler: Since/for	The water has been boiling for some minutes.	Correct (difficult level)
31	Filler: present simple	The girl always hurries to school.	Correct (easy level)
32	Filler: present continuous	Julia is not sitting next to David right now.	Correct (easy level)

Note. The first three items are practice sentences.

2.2.1. Photos used in the EIT (version B):

1-1

1-2

2-1

2-2

3-1

3-2

4-1

4-2

5-1

5-2

6-1

6-2

7-1

7-2

8-1

8-2

9-1

9-2

10-1

10-2

11-1

11-2

12-1

12-2

13-1

13-2

14-1

14-2

15-1

15-2

16-1

16-2

17-1

17-2

18-1

18-2

19-1

19-2

20-1

20-2

21-1

21-2

22-1

22-2

23-1

23-2

24-1

24-2

25-1

25-2

26-1

26-2

27-1

27-2

28-1

28-2

29-1

29-2

30-1

30-2

31-1

31-2

32-1

32-2

2.3. Timed Grammaticality judgement task (Version A)

	Structure	Stimulus	Time limit (Seconds)	Error type
I	NA	Never seen have I anything like this.	3	Placing V after NA
II	NA	Seldom does Smith ask for help in a polite way.	4	Correct
III	TQ	Tom refused to discuss the matter publicly, he didn't?	4	Negative tag question: AUX-S inversion
IV	TQ	You sometimes sleep on you back, aren't you?	3.5	Correct
V	CON	My brother had took the map, he would have arrived earlier.	5	First clause: Swapping AUX with S; past form of the V
VI	CON	Had I invited John, my mum would have brought Smith.	5	Correct
1	TQ	The film finished much earlier last night, did it?	6.155	Affirmative-Affirmative
2	TQ	Those apples shouldn't be ready now, shouldn't they?	4.635	Negative-Negative
3	TQ	Your roommate will not be late for the dinner party, he will?	5.090	Affirmative tag question: AUX-S inversion
4	TQ	The trainer has broken his arm recently, he hasn't?	4.954	Negative tag question: AUX-S inversion
5	TQ	That woman was a professor at Bristol University, wasn't she?	4.937	Correct
6	TQ	The daughters normally cook during the week, don't they?	4.715	Correct
7	TQ	The students are waiting at the train station now, aren't they?	5.623	Correct
8	TQ	Your father can't help you with your tuition fees, can he?	4.870	Correct
9	NA	Seldom my grandmother would use the lift in the shopping mall.	5.714	Swapping AUX with S
10	NA	Rarely the students have received feedback on their projects.	5.256	Swapping AUX with S

11	NA	Did rarely Helen clean the house on workdays.	3.738	Swapping NA with AUX
12	NA	Does seldom my daughter admit his errors in class.	3.932	Swapping NA with AUX
13	NA	Rarely will Alex study so hard for the final English test.	5.898	Correct
14	NA	Seldom does my friend have flu in the British winter.	6.001	Correct
15	NA	Seldom will David see such a beautiful display of flowers.	6.669	Correct
16	NA	Rarely has my grandfather laughed at my jokes.	4.665	Correct
17	CON	Had the adviser known all the facts, he would finish the report on time.	6.138	Second clause: Would + V
18	CON	Had John found their address, he would send their invitation.	6.324	Second clause: Would + V
19	CON	Had my cousins listened to me, they found the flat.	4.139	Second clause: Past form of the V
20	CON	Had the applicant interviewed better, he got the job.	4.496	Second clause: Past form of the V
21	CON	Had you answered the phone, I would have told you the news.	4.512	Correct
22	CON	Had Tom passed the maths exam, he would have gotten a better job.	5.884	Correct
23	CON	Had my mother cooked more, no-one would have gotten hungry.	6.330	Correct
24	CON	Had the car broken down, we would have called a taxi.	4.777	Correct
25	Filler: Relative clause	This is the friend which wife wants to get divorced.	5.679	Swapping 'whose' with 'which'(difficult level)
26	Filler: Since/for	We have been doing the housework from we arrived home.	6.568	Swapping 'since' with 'from' (difficult level)
27	Filler: Present simple	Michael go to the supermarket every morning.	4.093	Deleting -s (easy level)
28	Filler: Present continuous	The children watching a movie on TV right now.	6.115	Deleting 'are' (easy level)

29	Filler: Relative clause	The site where the party will take place is Southmead Park.	5.003	Correct (difficult level)
30	Filler: Since/for	The student has been working on the task for several days now.	4.552	Correct (difficult level)
31	Filler: Present simple	My child likes to watch cartoons every morning.	3.555	Correct (easy level)
32	Filler: Present continuous	The technicians are working on a project this year.	4.355	Correct (easy level)

Note. The first six sentences are practice sentences. The ungrammatical patterns for the NA and CON practice sentences differ from those used in the actual test sentences. I established the time limits for practice sentences arbitrarily; they were not based on native speakers' average response time.

2.4. Timed Grammaticality judgement task (Version B)

	Structure	Stimulus	Time limit (Seconds)	Error type
I	NA	Never seen have I anything like this.	3	Placing V after NA
II	NA	Seldom does Smith ask for help in a polite way.	4	Correct
III	TQ	Tom refused to discuss the matter publicly, he didn't?	4	Negative tag question: AUX-S inversion
IV	TQ	You sometimes sleep on you back, aren't you?	3.5	Correct
V	CON	My brother had took the map, he would have arrived earlier.	5	First clause: Swapping AUX with S; past form of the V
VI	CON	Had I invited John, my mum would have brought Smith.	5	Correct
1	TQ	The receptionist left work early on Wednesday, did he?	6.155	Affirmative-Affirmative

2	TQ	Those valuable paintings shouldn't be for sale, shouldn't they?	4.635	Negative-Negative
3	TQ	You will not present your MA thesis tomorrow, you will?	5.090	Affirmative tag question: AUX-S inversion
4	TQ	The wedding dress has arrived safely, it hasn't?	4.954	Negative tag question: AUX-S inversion
5	TQ	The house next to yours was built in the 1920s, wasn't it?	4.937	Correct
6	TQ	The twins don't like to sing to their parents, do they?	4.715	Correct
7	TQ	My roommates are far too busy to help, aren't they?	5.623	Correct
8	TQ	Your teacher can't help you with your art assignment, can he?	4.870	Correct
9	NA	Seldom the postman would deliver parcels before noon.	5.714	Swapping AUX with S
10	NA	Rarely the visitors have seen such stunning designs in fashion.	5.256	Swapping AXU with S
11	NA	Did rarely Lucy get up before 10 o'clock at the weekend.	3.738	Swapping NA with AUX
12	NA	Does seldom my wife make her own bread.	3.932	Swapping NA with AUX
13	NA	Seldom will you find a medical translator in hospitals.	5.898	Correct
14	NA	Seldom does the teacher let us leave the class early.	6.001	Correct
15	NA	Rarely have the patients complained about the poor care.	6.669	Correct
16	NA	Rarely do my grandparents go away on holiday.	4.665	Correct
17	CON	Had Tom studied well, he would answer the question better.	6.138	Second clause: Would + V
18	CON	Had David trained more, he would win the race.	6.324	Second clause: Would + V
19	CON	Had you left home early, you avoided the traffic jam.	4.139	Second clause: Past form of the V

20	CON	Had the boy finished studying before noon, he went out for lunch.	4.496	Second clause: Past form of the V
21	CON	Had I forgotten to bring my wallet, Martin would have paid for me.	4.512	Correct
22	CON	Had my computer crashed, I wouldn't have sent the email to you.	5.884	Correct
23	CON	Had I reserved a ticket early, I would have gone to the opera.	6.330	Correct
24	CON	Had my parents met earlier, I would have had more siblings.	4.777	Correct
25	Filler: Relative clause	I talked to the boy which car had broken down last week.	5.679	Swapping 'whose' with 'which'(difficult level)
26	Filler: Since/for	Liza has been writing for a local magazine from she graduated.	6.568	Swapping 'since' with 'from' (difficult level)
27	Filler: Present simple	My mother always close the windows at night.	4.093	Deleting -s (easy level)
28	Filler: Present continuous	The researchers designing an experiment at present.	6.115	Deleting 'are' (easy level)
29	Filler: Relative clause	My children saw the hall where the festival will take place.	5.003	Correct (difficult level)
30	Filler: Since/for	My brother has taught at this school for a long time.	4.552	Correct (difficult level)
31	Filler: Present simple	His wife often loves to watch comedy movies.	3.555	Correct (easy level)
32	Filler: Present continuous	The children are playing computer games at the moment.	4.3557	Correct (easy level)

Note. The first six sentences are practice sentences. The ungrammatical patterns for the NA and CON practice sentences differ from those used in the actual test sentences. I established the time limits for practice sentences arbitrarily; they were not based on native speakers' average response time.