

THE ROLES OF OUTPUT-INDUCED NOTICING IN L2 ACQUISITION: A PROCESS- AND
PRODUCT-ORIENTED STUDY THROUGH EYE-TRACKING

By

Kiyotaka Suga

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2024

ABSTRACT

Since Swain's (1985) Output Hypothesis, producing output in second language (L2) has been assumed to be a crucial cognitive process that promotes L2 acquisition, by actively facilitating various cognitive processes (e.g., noticing, hypothesis testing, conscious reflections of own language use, and automatization of the linguistic knowledge) (de Bot, 1998, Gass, 1997; Izumi, 2003; Muranoi, 2007a; Swain, 1985, 2005). In particular, the noticing function of output has been widely accepted as a cognitive rationale of L2 learning in Second Language Acquisition (SLA)/Instructed SLA (ISLA). Despite the commonly accepted Swain's Output Hypothesis and the noticing-inducing function of output, previous empirical studies that investigated whether and how producing L2 output could induce various types of learner-noticing and impact overall grammar learning gains have reported mixed results. Therefore, this study investigated how producing L2 output could induce learner noticing in a subsequent input (i.e., output-induced noticing) and contribute to the learning of the English past counterfactual/hypothetical conditional through a hybrid design of both product- and process-oriented research approaches.

The participants of the study were 117 international undergraduate and graduate students in the U.S. They were assigned to one of the three conditions (Oral Output, Written Output, and Input-Only Comparison Groups) and then engaged in each treatment task, respectively. During the treatment sessions, all the learners followed the same four instructional steps: (1) listened to an oral introduction that provided background knowledge of a reading text; (2) read a short text (First Input); (3) engaged in each different treatment task (Oral Output, Written Output, or Aural Input); and (4) read the same text again (i.e., subsequent input) (Second Input). Both output groups engaged in the same text-reconstruction task, in which they were asked to reconstruct the text that they had just comprehended as accurately as possible using descriptive picture cues.

However, the modality of the reconstruction was different (an oral or a written mode). The Input-Only Group listened to a text narration while watching the same descriptive picture cues. During the first and second input, learners' noticing behaviors on the target form were measured through an online, objective measure (i.e., eye-tracking) with two different levels of processing (i.e., the early [first-pass reading time, FPRT and the late measures [re-reading time, RRT]). In addition to these two process-oriented measures, the learners' overall learning gains were also assessed through a written-picture description test (WPDT) and an oral elicited imitation test (OEIT) in a pretest and posttest design as product-oriented measures.

Overall, the results of the study revealed that producing L2 output as a form of text-reconstruction induced learner noticing, which was evidenced by both Oral and Written Output groups' significantly increased eye-fixation duration to process the features of the target grammatical form in the subsequent input, whereas the Input Only group showed significantly decreased eye-fixation duration to process the form. The degree of output-induced noticing was moderated by the modality of output and the levels of the eye-tracking (FPRT or RRT). The results of the late measure of eye-tracking (i.e., RRT) indicated similar eye-fixation duration gains from the first reading to the second reading for both groups, but the results of the early measure (i.e., FPRT) showed a significantly higher FPRT for the Written Output group but not for the Oral Output group. As for the impact of L2 output on the acquisition of the target form, however, the grammar test results did not show any significant group differences, even though slight differences were indicated in the descriptive results. Therefore, the findings of the study did not indicate measurable effects of L2 output on grammar learning but the eye-tracking results demonstrated the detailed mechanisms of how the noticing-triggering function of output was promoted in the subsequent input after engaging in L2 output practice.

Copyright by
KIYOTAKA SUGA
2024

This dissertation is dedicated to Erie.

ACKNOWLEDGEMENTS

This dissertation would not have been completed without the help of a number of people. I would like to express my sincere gratitude to all of them who supported this project.

My deepest gratitude goes first and foremost to my mentor, Professor Shawn Loewen, for guiding me to accomplish my long journey of the dissertation project. Without his generous guidance and both intellectual and psychological support, I would not have been able to complete this project. Furthermore, I am deeply indebted to Professor Loewen for leading me into the field of Instructed Second Language Acquisition (ISLA). Since I attended his two-day intensive ISLA seminar at Temple University in Tokyo eight years ago, his passion and work kept inspiring me to further study and explore numerous topics in ISLA. Now, I am very glad that I followed him and came to study at Michigan State University. All the experiences that I had with Professor Loewen here are precious experiences for me.

I would also like to thank all the committee members, Dr. Aline Godfroid, Dr. Paula Winke, Dr. Koen Van Gorp, and Dr. Masatoshi Sato, for giving me numerous pieces of advice at each stage of working on the current dissertation project. I gained numerous insights for the design, analyses, and interpretations of the findings from their comments and suggestions. Without their support, I would not have been able to complete this dissertation project.

This research would not have been possible without the cooperation of the participants. I would like to express my deep appreciation to all of them for eagerly participating in this study while managing their coursework at the university.

I would also like to express my gratitude for the Language Learning Dissertation Grant and the dissertation grants from the College of Arts and Letters and Second Language Studies.

Lastly, I would like to express my genuine gratitude to my mother for her wholehearted support and encouragement.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: LITERATURE REVIEW	6
CHAPTER 3: METHOD.....	41
CHAPTER 4: RESULTS.....	78
CHAPTER 5: DISCUSSIONS.....	102
CHAPTER 6: CONCLUSION AND LIMITATIONS.....	119
REFERENCES.....	127
APPENDIX A: WEBB, SASAO, AND OLIVER’S (2017) UPDATED VOCABULARY LEVELS TEST	137
APPENDIX B: RECRUITING FLYER	146
APPENDIX C: A PASSAGE USED FOR THE PRACTICE READING	147
APPENDIX D: SLIDES FOR THE ORAL INTRODUCTION FOR THE READING TEXT.....	148
APPENDIX E: NARRATION SCRIPT FOR THE ORAL INTRODUCTION FOR THE READING TEXT	154
APPENDIX F: READING TEXTS	158
APPENDIX G: ALL THE EXEMPLAR SENTENCES AND AOIS	160
APPENDIX H: WRITTEN PICTURE-CUED DESCRIPTION TEST.....	162
APPENDIX I: ALL THE OEIT AND THE DETAILS FOR EACH TEST ITEM.....	178
APPENDIX J: AN EXAMPLE DISPLAY FO THE OEIT TEST FORMAT	180
APPENDIX K: ASSUMPTION CHECKING FOR MIXED-DESIGNED ANOVA FOR THE GRAMMAR DEVELOPMENTAL TESTS (OEIT & WPDT)	181

CHAPTER 1: INTRODUCTION

Background of the Study

Producing output in second language (L2) plays a crucial role in promoting L2 acquisition by inducing different types and levels of noticing, pushing learners to engage in hypothesis formulation/testing and conscious reflections of their own language use, and automatizing their linguistic knowledge, rather than just being an overt manifestation of the end product of what learners have already acquired (e.g., Gass, 1997; Izumi, 2003; Leow, 2015; Muranoi 2007a; Swain, 1985, 2005; Zalbidea, 2021). Previous output studies indicated that the act of producing output induces learner noticing particularly in subsequent input-processing opportunities because learners try to search for more accurate, precise, and appropriate linguistic information while processing the subsequent input based on the mismatches between their current linguistic knowledge and what they wanted to express (i.e., noticing the holes/gaps in one's ability), which they found during their prior output production (Izumi, 2002; Swain, 1995, 2005; Zalbidea, 2021). This type of internal priming of learner noticing after producing output has been claimed as one of the theoretical rationales for the beneficial role of output in L2 acquisition (Gass, 1997; Izumi, 2003; Leow, 2015; Muranoi, 2007a). However, previous studies that investigated whether and to what extent L2 output promotes learner noticing as well as L2 learning focused primarily on the product of L2 learning employing the pretest and posttest design (i.e., a product-oriented approach). Along with exploring the amount of noticing and learning through this type of product-oriented approach, closely examining how L2 learners direct their attention to and process certain target linguistic forms while they are engaging in treatment tasks (i.e., process-oriented approaches) has been attracting increasing attention from researchers in the field of ISLA (Ellis, 2001; de Graaff & Housen, 2009; Hanaoka & Izumi,

2021; Leow, 2015; Gilabert et al., 2016). Particularly, the effect of output on learner noticing has been explored both quantitatively and qualitatively through various online and offline measuring techniques, such as note-taking (e.g., Hanaoka, 2007; Izumi, 2002; Leeser, 2008), underlining (e.g., Ghari & Moizadeh, 2011; Izumi & Bigelow, 2000; Russell, 2014; Song & Suh, 2008; Uggen, 2012), retrospective questionnaire (e.g., Izumi & Izumi, 2004), and stimulated recall (e.g., Uggen, 2012, Zalbidea, 2021). Some studies have shown supportive evidence for both the noticing-inducing function of output and the learning of some specific linguistic forms (e.g., Russell, 2014; Uggen, 2021; Zalbidea, 2021) or at least for the learning of a certain grammatical form (e.g., Izumi, 2002), whereas other studies failed to show consistent and convincing evidence of these beneficial roles of L2 output due to inconsistent operationalization of output and output-induced noticing and various methodological limitations. These inconsistent results and methodological limitations called for further accumulation of empirical findings with a process-oriented approach using consistent operationalizations of output and noticing as well as measures that can accurately examine how L2 output plays roles and enhances learner noticing in the overall processes of L2 acquisition (Leow, 2015).

One of the biggest limitations of previous output and noticing studies was a primary reliance on relatively indirect measures of learner noticing, such as note-taking, underlining, or retrospective verbalizations. Accordingly, it is possible that learner noticing was not fully captured with these indirect noticing measures in previous studies (Godfroid, 2019, 2020; Godfroid & Uggen, 2013; Winke, 2013). In other words, what learners failed to report or verbalize could not be measured through these indirect techniques. Therefore, it was crucial to further examine the roles of L2 output in relation to learner noticing as well as the overall processes of L2 acquisition.

From the pedagogical perspective, most L2 teachers may believe intuitively that engaging in output practice in L2 classrooms is crucial for learners to develop their L2 knowledge. As discussed above concerning the measuring issues of learner noticing and the lack of focus on the modality difference (oral or written output) of L2 output, revisiting Swain's Output Hypothesis and then further specifying the detailed mechanisms of the noticing-triggering function of output with a sensitive, online objective measure of learner noticing through eye-tracking method would provide L2 teachers with their instructional foundations that could allow them to accurately and critically evaluate various output-based classroom instructional strategies and also incorporate theoretically- and empirically-based output instruction into their daily pedagogical practices. As for the modality difference, the findings also allow teachers to make pedagogical decisions on which modality to use for what purposes within their limited class time.

Aims of the Study

Based on these inconsistent results and the limitations of the previous output and noticing studies, the primary purpose of the present study was to address the theoretical gaps and empirical issues on the roles of output in L2 grammar acquisition specifically focusing on output-induced noticing in subsequent input, which had been claimed to be enhanced through learners' prior output production, and its contribution to the overall learning of L2 grammar. To closely examine this issue, this study employed a hybrid design of both product- and process-oriented research approaches using an online objective measure (i.e., eye-tracking) along with grammar development tests with the pretest and posttest design. More specifically, the present study addressed whether and to what extent producing L2 output could enhance learner noticing of the target linguistic form (i.e., the English past hypothetical conditional) in the subsequent input and promote the learning of the grammatical form. Additionally, this study also examined

the potential impact of output modality (oral or written output), which has not been systematically and consistently examined in previous output and noticing studies. Furthermore, the associations between learner-noticing induced by L2 output production and the overall L2 grammar acquisition were also examined by using measures that can shed light on both the process and the product of L2 grammar acquisition.

The findings of this study can provide both theoretical and pedagogical implications. Since the eye-tracking measure examined the learners' detailed noticing behaviors, which had not been objectively observable in previous output and noticing studies, the findings of the study contribute to clarifying the fine-grained mechanisms of Swain's Output Hypothesis and the noticing-inducing function of output with empirical evidence of learners' internal cognitive processes of L2 acquisition. Additionally, the study can also provide methodological implications for future studies to re-examine the mixed results reported in previous output and noticing studies using eye-tracking measures and stimulated recalls. In this sense, this study contributed to the research field theoretically and methodologically.

Organization of the Study

This dissertation consists of six chapters. Following this introduction chapter, Chapter 2 reviews previous literature on the role of output in L2 acquisition, the relationships between L2 output, noticing, and grammar acquisition, empirical findings and issues from previous output/noticing studies, effects of the differences in L2 output on learner noticing and grammar learning, and measuring issues for L2 noticing. All these issues motivated this dissertation study to address the research questions. In Chapter 3, the methodological procedures of the present study are presented including the whole research design, details of the participants, the target forms, procedures for the instructional treatments, measuring and testing instruments, scoring

procedures, and analyses of the data. In Chapter 4, the results of the process-oriented measures (i.e., eye-tracking and stimulated recalls) and the product-oriented measures (i.e., the OEIT and the WPDT) are given. Descriptive statistics, their visual representations, and the results of all the quantitative analyses are presented. Discussions on the results presented in Chapter 4 are provided in Chapter 5. The last chapter (Chapter 6) summarizes the major findings and describes the limitations and pedagogical implications of the present study.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the literature on the main areas of interest of this study. First, an overview and the primary goal of the field of Instructed SLA (ISLA) are introduced followed by the overview of the overall cognitive processes of L2 acquisition from input to output by referring to the integrated model of SLA (Gass, 1988, 1997; Leow, 2015). Based on these overviews of the basic framework of the study and the cognitive rationale for both the processes and the products of L2 learning, the roles of L2 output are reviewed based on Swain's (1985, 1995, 1998, 2005) Output Hypothesis and the four major functions output (i.e., *automatization*, *hypothesis testing*, *metalinguistic*, and *noticing triggering*). Since learner noticing is the most essential theoretical foundation of the Output Hypothesis, the theoretical background of Schmidt's (1990, 1993, 1994, 1995, 2001, 2012) Noticing Hypothesis is introduced and the detailed mechanisms of the noticing-triggering function of output in L2 grammar learning are clarified. While reviewing the detailed mechanisms of the cognitive processes enhanced by the noticing-triggering function of output, the operationalization of output-induced noticing that the current study focused on is provided (i.e., noticing a form-meaning-function relationship and noticing the gap between interlanguage [IL] and target language [TL] that are induced while processing subsequent input after producing output). After clarifying various types of output-induced noticing, empirical findings from previous output and noticing studies are reviewed and research gaps and methodological issues are identified and discussed. Since measuring learner noticing is a major issue that has contributed to the mixed results of previous output and noticing studies, the strengths and weaknesses of various types of noticing measures that have been conducted in previous studies are reviewed and the importance of employing more sensitive, online objective measures of learner noticing (e.g., eye-tracking) together with offline subjective

verbal reports is discussed. Finally, the potential roles of output modality differences (either oral or written output modality) on learner noticing and the overall cognitive processes of L2 grammar learning are reviewed to introduce the primary aims and designs of the current study. Based on the literature reviews and discussions provided in this chapter, five research questions and the corresponding hypotheses are proposed at the end of this chapter.

Instructed Second Language Acquisition and the Integrated Model of L2 Acquisition

One of the engaging questions for L2 researchers and teachers is how L2 instruction can be optimized to best facilitate L2 learners' acquisition of various linguistic features (Spada & Lightbown, 2008; Loewen, 2015). Particularly, how L2 grammar instruction can facilitate adult L2 learners' learning of late-acquired, challenging grammatical forms is a pressing issue for many L2 teachers (Celce-Murcia & Larsen-Freeman, 2016). Such empirical and pedagogical questions are important domains of inquiry in the field of Instructed Second Language Acquisition (ISLA). Based on Loewen's (2015) definition of ISLA, it can be defined as "a theoretically and empirically based field of academic inquiry that aims to understand how the systematic manipulation of the mechanisms of learning and/or the conditions under which they occur enable or facilitate the development and acquisition of language other than one's first" (p. 2). Therefore, findings from ISLA research can provide implications that allow L2 teachers to accurately and critically evaluate various instructional strategies and also make pedagogical decisions to incorporate theoretically- and empirically-based instructional practices into their daily teaching.

Among various aspects of ISLA, one of the major approaches that has been focused on, in particular, is cognitive approaches to ISLA, which has examined mental processes that explain how L2 knowledge is represented and acquired through instruction based on theories and

mechanisms of cognitive and psycholinguistic processes of L2 acquisition (Ellis, 2008; Loewen, 2015; Leow, 2019). To achieve the primary aim described in Loewen's (2015) definition, especially within the domains of cognitive approaches to ISLA, a useful framework of L2 learning processes is Gass' (1988, 1997) integrated cognitive model of L2 acquisition (see Figure 2.1; also see Leow, 2015 for the updated model focusing on both the processes and the products of L2 learning). This model has been used as an important psycholinguistic rationale in numerous ISLA studies and depicts a fine-grained framework of how L2 learners' interlanguage system develops by converting input that learners receive to output that learners can produce (e.g., Ellis, 2008; Gass, 1988, 1997, 2013; Izumi, 2003, 2013, Leow, 2015; VanPatten, 1996). This cognitive model consists of five major stages of L2 acquisition processes: apperceived input (noticed input), comprehended input, intake, integration, and output (Gass, 1988, 1997; see Figure 2.1).

As shown in Figure 2.1, the starting point of this model is apperceived input (or noticed input), which is the type of input that learners pay attention to out of the entire L2 linguistic information that learners are exposed to (i.e., ambient input). To proceed further to the following stage of the acquisition processes, learners need to direct their attention to certain aspects of input selectively from the whole exposure of ambient input. Thus, this type of apperception (or the selective attention to certain linguistic features) is a priming device for further linguistic analyses (or deeper processes), which, according to Gass (1997), can be influenced by learners' previous experiences as well as their existing knowledge. The amount of apperceived input can also be influenced by various other factors (e.g., frequency of input, learners' affective state, and the type of their L2 learning/instructional experiences, etc.).

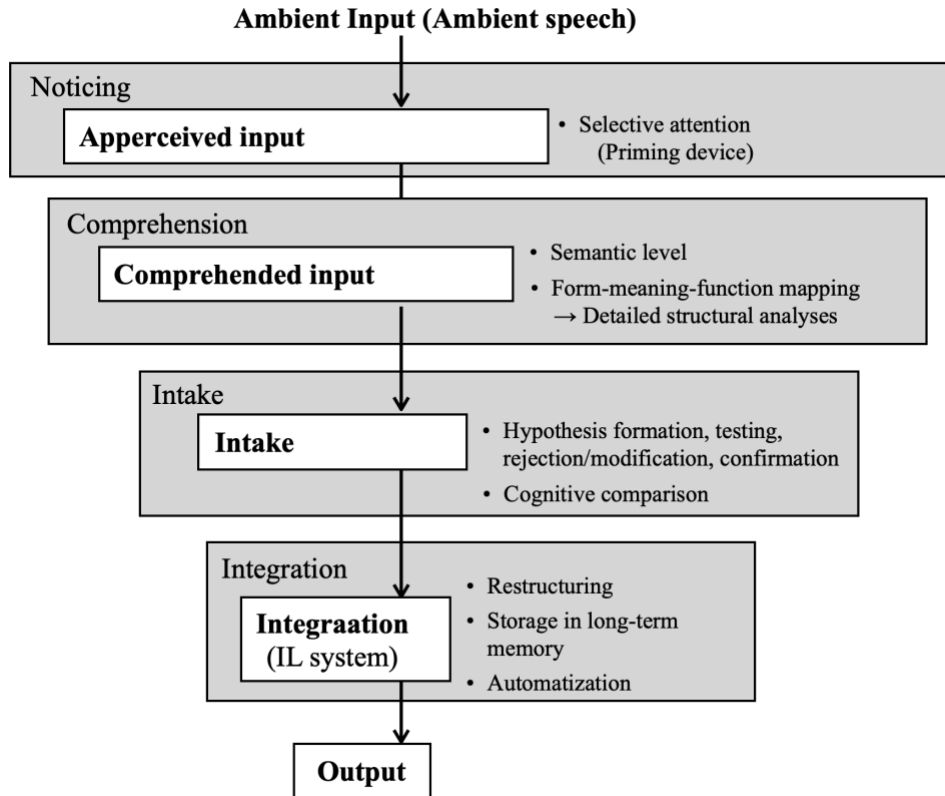


Figure 2.1. Gass' integrated model of L2 acquisition (Adopted from Gass, 1988, p. 200)

*IL = Interlanguage

The next stage of the conversion in the model is comprehended input. The definition of comprehension here is not exactly the same as the one proposed by Krashen (1982, 1985) in his comprehensible input, which was merely defined depending on whether the meaning of the message was comprehended or not. The comprehended input in this model consists of different levels of comprehension ranging from semantic to detailed structural analyses. To reach the latter level of linguistic analyses, comprehended linguistic information needs to be mapped onto the form, meaning, and function of that particular linguistic feature(s), which further facilitates hypothesis formation(s) in the next stage (i.e., intake).

The third stage of the acquisition processes is intake, which is “the process of assimilating linguistic material; it refers to the mental activity that mediates input and grammars” through hypothesis formation, testing, rejection/modification, and/or confirmation (Gass, 1997,

p. 5). These assimilation processes are also called selective processing by matching the new linguistic information to the learner's prior or existing knowledge (i.e., cognitive comparison[s]).

The information processed in intake can be integrated into learners' long-term memory (i.e., interlanguage system). This stage of the process is called Integration, in which the reorganization of learners' internal knowledge system is often required. This reorganization process is referred to as restructuring (Maclaughlin, 1990). If the language information has already been integrated into the learner's interlanguage system when they receive it, the information contributes to the process of rule-reconfirmation and hypothesis strengthening (i.e., automatization).

The final stage of this model is output, which is often called the overt manifestation of the entire acquisition processes. However, Gass (1997) claims that output is not just the final product of L2 acquisition but rather is another important stage that actively facilitates further L2 acquisition, creating feedback loops back into the prior stages of the acquisition processes (also see Boers, 2021; Izumi 2003, 2013; Leow, 2015; see further discussions on specific functions of output in the section of Swain's Output Hypothesis in this chapter).

As described above, Gass' (1997) integrated model of L2 acquisition provides psycholinguistic rationale and important implications for L2 learning with its fine-grained demonstrations of how L2 knowledge is processed and is eventually acquired through various dynamic cognitive processes from input to output. First, the model emphasizes the importance of receiving a large amount of input, which is an indispensable prerequisite for L2 acquisition. Even if the main purpose of learning is to develop learners' output, learners still need to first receive and process input to facilitate the overall acquisition processes. Another important implication from this model is that only some portions of linguistic information can be acquired, while other

information is filtered out at various stages of the acquisition processes. Therefore, not only the quantity of input but also the quality of language processing or the depth of processing at each stage of acquisition processes in the model are crucial to converting more linguistic input into output. In relation to the purpose of the current study, the most important implication from the model is the demonstration of how producing L2 output itself can also play an active role in further promoting L2 acquisition by facilitating other stages of the acquisition processes from input to output and *again back to input* while it is also represented as a final product of the overall acquisition processes in Gass' (1988, 1997) integrated model of L2 acquisition. In other words, engaging in L2 output is one way to internally promote L2 learning (i.e., internal priming) based on the cognitive feedback loop back from output to input.

Roles of Output in L2 learning

As discussed in the previous section by referring to Gass' (1988, 1997) integrated model of L2 acquisition, as well as comprehending input, which is the most essential driving force that promotes L2 acquisition, producing output is also a dynamic cognitive act that plays multiple roles in the overall SLA processes, rather than just an overt manifestation of the end product of L2 acquisition (de Bot, 1998, Gass, 1997; Izumi, 2003; Muranoi, 2007a; Swain, 1985, 1995, 1998, 2005). After observing Canadian French-immersion students' persistent failure to acquire certain aspects of grammatical forms despite many years of receiving comprehensible input in their programs, Swain (1985) proposed the Output Hypothesis and addressed the issues of Krashen's (1985) Input Hypothesis which claimed that language acquisition occurs only through understanding comprehensible input at $i+1$ (i.e., slightly beyond the current level of L2 competence) and such input is necessary and sufficient for the development of L2 knowledge. According to Krashen's (1985) Input Hypothesis, Canadian French immersion is supposed to be

the optimal condition for L2 learning. However, Swain's observation revealed that while the immersion students achieved excellent comprehension skills, they still failed to achieve target-like performances in some morphosyntactic aspects. Therefore, she argued that L2 learners need not only comprehensible input but also comprehensible output, which is a kind of output that L2 learners are cognitively and situationally pushed to produce to make their utterances more accurate, precise, and appropriate. Hence, engaging in pushed output "may force learners to move from semantic processing to syntactic processing" (Swain, 1985, p.249). In particular, the opportunities to engage in pushed output have been identified to have four major functions: (1) providing opportunities to practice and automatize existing linguistic knowledge (i.e., *the automatization function*) (also see de Bot, 1996), (2) enabling learners to test their hypotheses about their developing L2 knowledge (i.e., *the hypothesis testing function*), (3) providing opportunities to consciously reflect on their own L2 use and their L2 knowledge as they engage in a dialogue with others, which raises their metalinguistic awareness (i.e., *the metalinguistic function*), and (4) triggering various types of noticing (e.g., noticing the gap and holes) (i.e., *the noticing(-triggering) function*). All these major functions of output have been predicted to contribute to the development of L2 knowledge (de Bot, 1996, Gass, 1997; Izumi, 2003; Muranoi, 2007a; Swain, 1985, 1995, 1998, 2005).

Roles of Learner Noticing in L2 Learning

Out of these four major functions of output, the noticing-triggering function of output has been considered as one of the most important theoretical rationales for engaging in L2 output in the field of ISLA (Loewen, 2015). Regarding the noticing-triggering function of output, Schmidt's Noticing Hypothesis is the key theoretical foundation for this function (Schmidt, 1990, 1993, 1994, 1995, 2001, 2012). Before reviewing the detailed mechanisms of the noticing-

triggering function of output, the roles of learner noticing in L2 learning are reviewed in this section.

First of all, noticing is only the first step of the overall L2 acquisition processes but is an essential prerequisite for converting linguistic information included in input into intake (Izumi, 2013; Schmidt, 1990, 2001, 2012) (also see Figure 2.1). Schmidt's Noticing Hypothesis proposed that to learn any linguistic features, L2 learners must attend to and notice the specific linguistic feature at least at a low level of awareness (Schmidt, 1990, 1993, 1994, 1995, 2001, 2012). In other words, without any focal attention or a low level of awareness of the target linguistic forms, L2 learning cannot take place (Schmidt, 1995). In his later publications, Schmidt further claimed that "more noticing leads to more learning" (Schmidt, 1994, p. 129) because learner noticing is a cognitive process that interacts with various learner-internal and external factors, such as motivation and other individual difference variables (e.g., aptitude, learning experiences, among others) (Schmidt, 2001, 2012).

Since Schmidt (1990) postulated his seminal Noticing Hypothesis, the notion of noticing has become an essential, underlying construct that arguably promotes L2 learning and has contributed to the development of theories of SLA/ISLA. However, the definition of noticing in Schmidt's sense is not as clear as he claimed, thereby having attracted debates and controversies over the operationalization of noticing and its roles in L2 learning among SLA researchers (Robinson, 1995; Schmidt, 1990, 2001, 2012; Tomlin & Villa, 1994).

In Schmidt's operationalization of noticing, both attention and awareness are necessary for L2 learning (Schmidt, 1990, 2001). Particularly, he claimed that awareness at the level of understanding, which is a relatively high level of awareness, enables learners to engage in deeper processing of the linguistic information and increases the likelihood of the information being

processed in the subsequent stages of acquisition (e.g., deeper analysis of the linguistic information, comparing, hypothesis testing, and restructuring) rather than just processing the information for mere comprehension- or communication-purposes. Concerning the necessity of awareness, however, Tomlin and Villa (1994) described the following three components of attention: (1) alertness (i.e., an overall readiness to deal with incoming stimuli), (2) orientation (i.e., the direction of attentional resources), and (3) detection (i.e., cognitive registration of stimuli), and then posited that only the last component of attention (i.e., detection) is a necessary condition for learning to take place. According to their model of input processing, awareness is not a necessary condition for L2 learning, and thus their claims partially contradicted Schmidt's postulation of the Noticing Hypothesis. As a reconciliation of these opposing views, Robinson (1995) defined the notion of noticing as "detection plus rehearsal in short-term memory" (p. 296). In his model, although detection is an important first step, further processes accompanied by awareness at the level of noticing are necessary for learning to take place.

While the operationalization of noticing (i.e., the amount of attentional resource and the degree of conscious awareness that is required for L2 learning) is one of the most abstract, controversial constructs in SLA/ISLA theories (e.g., Hama & Leow, 2010; Leow, 2000; Leung & Williams, 2011; Robinson, 1995; Schmidt, 1990; Tomlin & Villa, 1994; Williams, 2005), the main claims of Schmidt's Noticing Hypothesis have been widely accepted as one of the most fundamental underpinnings that account for the processes of L2 acquisition (Leow, 2015; Philip, 2013). Therefore, it is crucial for L2 noticing studies to carefully operationalize the construct of noticing that the researcher is focusing on and then examine how and what kind of noticing (e.g., L2 learners' attentional allocations and the degree of their conscious awareness) can be enhanced and serve overall L2 learning while learners are actually engaging in L2 learning using a

process-oriented research design (see the section of Measuring Issues of Learner Noticing in this chapter for more detailed discussions on process-oriented research).

Detailed Mechanisms of the Noticing-Triggering Function of Output

As Schmidt's Noticing Hypothesis proposed, learner noticing plays essential roles in promoting L2 learning (Schmidt, 1990, 1993, 1994, 1995, 2001, 2012). In relation to Swain's Output Hypothesis, the noticing-triggering function of output induces learners' noticing of holes in their interlanguage or gaps between what they want to say and what they can say, thereby enabling them to realize the possibilities and limitations of their linguistic knowledge (Swain, 1995). This type of realization serves as an internal priming device for consciousness-raising, which increases the learners' sensitivity toward the problematic forms that have not yet been fully integrated into their interlanguage system in the subsequent input (Gass, 1997; Izumi, 2003).

Swain's (1985, 1995, 1998, 2005) Output Hypothesis provided the mechanisms of the noticing-triggering function of output by referring primarily to two different types of noticing (i.e., noticing the gaps and holes) as described above. However, a closer re-evaluation of the noticing-triggering function of output in relation to the integrated model of SLA (see Gass, 1988, 1997; Izumi, 2003; Leow, 2015) is necessary and crucial to accurately understand: (1) what kind and level of noticing can be promoted through the act of producing output (i.e., the type of noticing enhanced by producing output), and (2) how the type of noticing induced by output production interacts with other stages of the overall L2 acquisition processes (i.e., the timing of noticing) (see Figure 2.2).

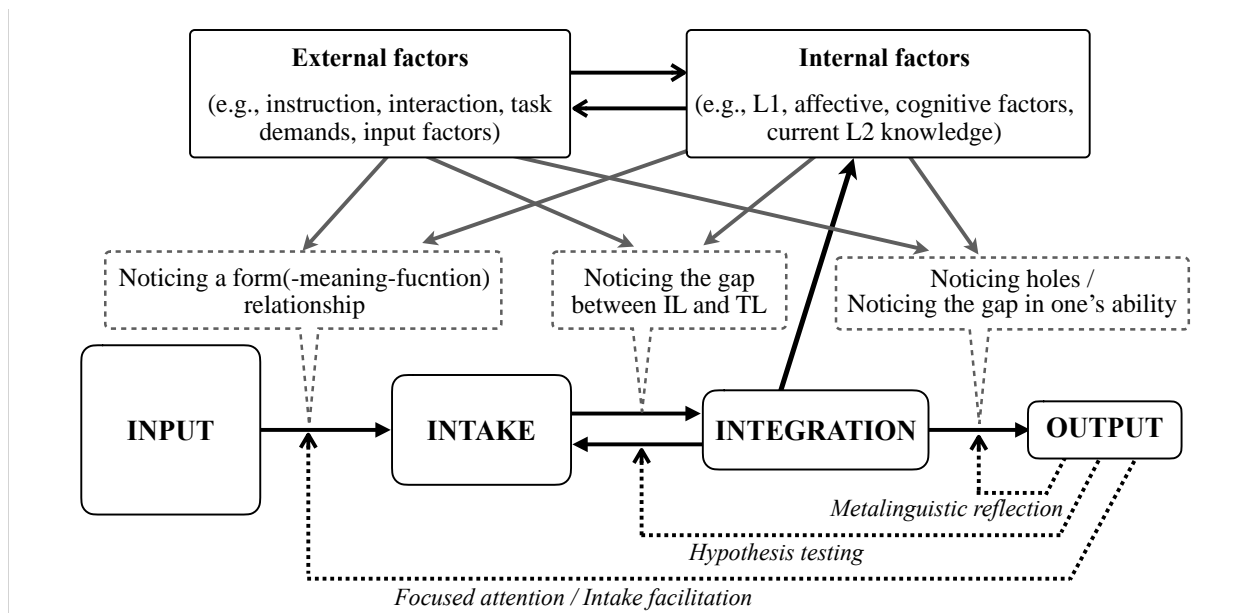


Figure 2.2. The types and the timing of noticing throughout the overall L2 acquisition processes (Adopted from Izumi, 2013, p. 41)

To address these two questions that were not specified in detail in Swain’s original output hypothesis, Izumi (2013) reviewed previous noticing studies and then further specified four different types of noticing based on the timing of noticing: (1) noticing a form-meaning-function relationship, (2) noticing the gap between interlanguage (IL) and target language (TL), (3) noticing holes in IL, and (4) noticing the gap in one’s ability (see Figure 2.2). The first type of noticing (i.e., noticing a form-meaning-function relationship) is the most basic concept of noticing proposed by Schmidt and Frota (1986). This type of noticing occurs when learners “notice how a particular form is used in the input they receive” (Izumi, 2013, p. 26). Noticing the gap between IL and TL is a more advanced type of noticing that facilitates a fine-tuning of the learner’s interlanguage by comparing “the difference between how the learner uses a language form and how a more proficient user uses it to convey the same idea” (Izumi, 2013, p. 26). The latter two types of noticing are relatively similar but noticing holes in IL is triggered when the learner notices a complete absence of certain grammatical forms or lexical items (holes) in his/her interlanguage whereas noticing the gap in one’s ability refers to a cognitive comparison

made by the learner internally between what he/she wants to say precisely or exactly and what he/she actually says based on his/her current IL knowledge. Realizing the existence of these holes and/or gaps is believed to encourage learners to direct their selective attention to relevant linguistic information in the subsequent input (Swain, 1998; Swain & Lapkin, 1995).

Although all these four types of noticing can be induced by producing output, each type of noticing is qualitatively different from each other, playing different roles at different stages of the acquisition processes (see Figure 2.2). The first two types of noticing (i.e., noticing a form-meaning-function relationship and noticing the gap between IL and TL) are triggered by producing output but occur while learners are processing input (i.e., feedback loop), instead of while learners are producing output or while learners are preparing for their production. In relation to the noticing-triggering function of output, these two types of output-induced noticing are promoted in the subsequent input-processing, which is after producing output. This type of noticing was the primary focus of the current study. On the other hand, the latter two types of noticing are the ones that are promoted while or right before learners are producing output. These qualitative differences in terms of how output induces noticing at which stages of the overall L2 acquisition processes need to be carefully specified to accurately examine and evaluate the noticing-triggering function of output (i.e., output-induced noticing) in L2 learning.

Effects of Output on Learner Noticing and L2 Learning

The noticing-triggering function of output in relation to the development of L2 knowledge has been investigated in previous studies (i.e., Ghari & Moinzadeh, 2011; Izumi, 2002; Izumi & Bigelow, 2000; Izumi & Izumi, 2004; Izumi et al., 1999; Kang, 2010; Leeser, 2008; Li & He, 2017; Muranoi, 2012; Russell, 2014; Shin, 2011; Song & Suh, 2008; Uggen, 2012; Zalbidea, 2021; also see Basterrechea et al., 2014; Li et al, 2016; Muranoi, 2007b;

Muraoka, 2006; Shintani, 2019, for studies that focused only on L2 learning gains through output tasks). Although the noticing-triggering function of output has been widely accepted in the field of SLA/ISLA, these previous empirical studies have reported mixed results in terms of output-induced noticing and the overall learning gains depending on various methodological variables, such as task types, target linguistic forms, the types of linguistic measures, the number of exposures to the target linguistic exemplars, the operationalization of noticing and its measures, and the modality of output during the treatment sessions (i.e., oral or written).

The seminal attempt to address this issue was made through a series of output and noticing studies conducted by Izumi and his colleagues (i.e., Izumi et al., 1999; Izumi & Bigelow, 2000; Izumi, 2002; Izumi & Izumi, 2004). Izumi et al. (1999) tested whether producing output promotes English as a second language (ESL) learners' noticing and the acquisition of the English past hypothetical/counterfactual conditional by comparing the experimental output condition (EG) and the comparison input condition (CG). The EG engaged in a series of two different output tasks in each of the two phases of the treatment sessions (Phase 1 and 2): (1) written text-reconstruction tasks and (2) guided essay-writing tasks. Instead of engaging in these output tasks, the CG worked on comprehension questions and an essay writing task that did not require the production of the target linguistic form in Phase 1 and 2, respectively. They investigated the amount of output-induced noticing in subsequent input (i.e., reading the original texts and the model essays, respectively, in each phase) through underlining and brief retrospective interviews by comparing those of the CG. The learning gains were measured through a grammaticality judgment test (GJT) and a picture-cued production test before and after each phase of the treatment tasks. Contrary to their original hypotheses, the results indicated no greater underlining scores for the EG than the CG. As for the learning gains, the EG

outperformed the CG on the production test only after Phase 2, which made the interpretation difficult because it was not possible to specify which output task or the combination of both tasks eventually contributed to the EG's observed learning gains after Phase 2. Furthermore, the interpretations of the noticing scores were also limited by the substantial individual variations of attentional allocations during the subsequent input processing among the participants.

Since their initial study did not show clear positive effects of output on learner noticing and L2 learning, Izumi and Bigelow (2000) replicated their original study by switching the order of the treatment task phases (i.e., Phase 1: the guided essay-writing task; Phase 2: the written text-reconstruction task) and changing one of the outcome measures from a written GJT to a multiple-choice fill-in-the-blank test. Again, no unique contribution of output on learner noticing and the acquisition of the target form (i.e., the English past hypothetical/counterfactual conditional) was found with great individual variations. Based on their careful reflections on their methodological design, they pointed out the importance of using less-free production tasks so that the learners can directly compare their previous problematic encodings with accurate, model encodings in the subsequent input. In their study, a text-reconstruction task was found to be less susceptible to individual variation of learner noticing in the subsequent input. They also suggested the incorporation of multiple measures in future research for achieving “much methodological triangulation (e.g., online measures, immediate retrospective report, task and test results)” (Izumi & Bigelow, 2000, p. 271).

After these two attempts of empirically investigating the noticing function of output, Izumi (2002) showed positive effects of output on the learning of L2 grammar. Contrary to the previous studies, this study targeted the acquisition of English relativization (i.e., the object-of-preposition [OPREP] type of English relativization). This time, he only employed a text-

reconstruction task with relatively short passages as the treatment task for the output conditions for the purpose of “maximizing the equivalence between the learners’ output and the target input” in the subsequent input-processing, which was assumed to efficiently promote noticing the gap between their IL and TL (p. 551). As a noticing measure, the learners’ note-taking behaviors in the subsequent input were used as a noticing measure instead of underlining. Although the consistent positive effects of output were evidenced by the results of the multiple development measures and the output groups’ reconstruction performances, the learners’ note-taking behaviors did not indicate any significant effects of output.

Since these three studies examined the effects of written output on L2 learning, Izumi and Izumi (2004) investigated the impact of oral output on L2 learning by targeting the same linguistic form as Izumi (2002) (i.e., the OPREP type of relative clauses). In this study, the output group engaged in a sentence-by-sentence picture-cued text-reconstruction task, in which participants were asked to listen to an aural model description of each picture in one sentence and reconstruct the sentence in the oral mode while looking at the picture. To avoid overtaxing their memory capacity for reconstruction, each picture-cued reconstruction was conducted one sentence at a time. To provide subsequent input, the aural model description was presented to them again after reconstructing the sentence. Contrary to their original expectations, the non-output group outperformed the output group in the posttest even though the results of the retrospective questionnaire indicated that the participants in the output group paid more attention to the target linguistic form in the subsequent input than did the non-output group. The authors claimed that their effort to reduce the processing load during the output task by having the learners reconstruct the text sentence-by-sentence allowed the participants to mechanically repeat the aurally presented sentence simply by memorizing it. Thus, it was likely that the participants

of the output group did not process the meaning of the text and thereby failed to engage in the genuine processes of production, which should start with having a meaningful message to tell at the stage of the conceptualizer (see Figure 2.3 and Levelt, 1989 for the production model; also see Kormos, 2006 for the bilingual production model). Regarding this methodological consideration, Izumi and Izumi (2004) argued that the output tasks need to be designed in a way that participants can engage in “grammatical encoding and that of monitoring to check the matching of the communicative intention generated in the conceptualizer with the output of the formulator” (p. 602) because these processes of output production serve as an internal priming device and then increase the learners’ sensitivity toward their problematic interlanguage forms in the subsequent input (see Figure 2.3; also see Izumi, 2003; Muranoi, 2007a). Therefore, the text length and the procedures of text-reconstruction tasks need to be considered carefully especially when the task is used in output and noticing studies. In other words, the text length in the text-reconstruction task should not be too long or too short. Either case may not provide learners with opportunities to promote their noticing and learning of the target linguistic form.

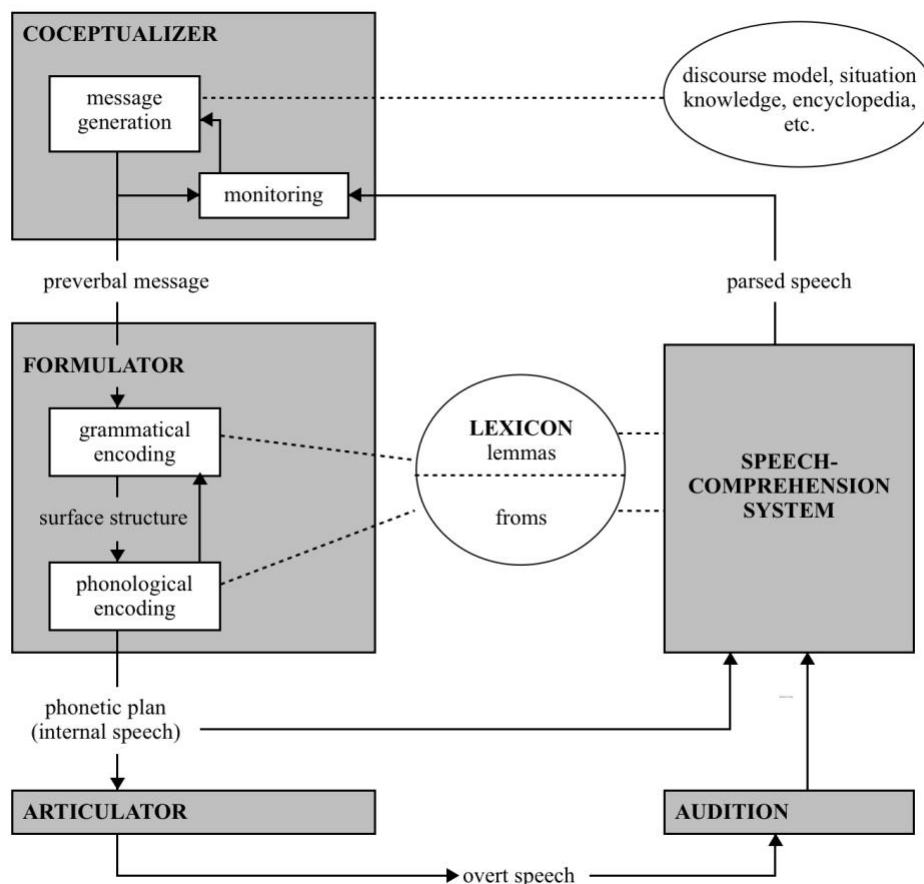


Figure 2.3. Levelt's Speech Production Model (Adopted from Levelt, 1989, p. 9)

After the series of Izumi's and his colleagues' attempts to clarify the roles of output in L2 learning, several replication and extension studies further investigated the effects of producing output on learner noticing and L2 learning. Song and Suh (2008) examined the relative effects of two different types of output tasks (i.e., a written text-reconstruction task and a picture-cued essay writing task) on learner noticing and the learning of the English hypothetical conditional by comparing these output conditions with the input-only condition. Forty-two adult Korean English as a foreign language (EFL) university students were assigned to one of these instructional conditions and engaged in each treatment task. The two output groups performed significantly better than the non-output group on the production test but not on the aural recognition test. The results of underlining scores that examined their rate of noticing indicated

no significant group differences in their gains underlying scores from the preceding input to the subsequent input even though both output groups showed a significantly higher rate of underlining than that of the non-output group. These contradicting results were attributed to the task instruction provided to the output group in advance. The authors argued that being aware of working on the output task after reading the first input (i.e., foreknowledge of the following task) changed their reading behaviors, resulting in their greater attention to their problematic linguistic forms even before producing output to better perform the following output task (also see Yoshimura, 2006; Russell, 2014). For example, Yoshimura (2006) investigated whether providing foreknowledge of the subsequent output tasks by letting learners know that they need to engage in output tasks after reading a text can induce more attention to linguistic form. The results indicated that providing foreknowledge of subsequent output tasks did change the learners' reading behaviors even before producing output, directing more attention to problematic linguistic form. Particularly, the learners who were told to engage in a text reconstruction output task after reading a text re-read the text more than those who were not informed to engage in the text-reconstruction after reading the text.

Since previous studies investigated output-induced noticing while learners were reading a text or processing written input as a form of subsequent input, Leiser (2008) examined the effects of output on learners' noticing of the Spanish past tense morphology (preterit/imperfect) in aural input. In this study, output was operationalized through the use of a dictogloss task, in which learners were asked to (1) listen to an aurally read text several times, (2) take notes based on their comprehension, (3) reconstruct the text in a written mode, and then (4) listen to the text again as a subsequent input. The non-output condition engaged in comprehension questions. The note-taking scores indicated more noticing for the output group than the non-output group but on

words (nouns) that were not related to the target linguistic form. As for the learning gains, no significant effects of output on the learning of the target form were reported in a writing test.

As a conceptual replication of Izumi (2002), Russell (2014) focused on the moderating effects of the target linguistic form on output-induced noticing and learning. Contrary to the original study, which targeted a relatively complex grammatical form (i.e., the OPREP type of relative clauses), this study focused on the Spanish future tense form, which is a more salient, less complex form. The results aligned with the ones of the original study, supporting the noticing function of output and the beneficial effects of L2 output on grammar learning. Although the note-taking measure in Izumi's (2002) study did not show clear evidence of output-induced noticing in the subsequent input, Russell (2014) attributed the learners' increased rate of underlining (i.e., noticing gains) to the property of the target linguistic form, which was a visually salient, meaning-bearing form. However, as a major limitation of the study, Russell (2014) claimed, "the noticing measure (underlining) may have been too coarse and likely did not measure all the noticing that took place" (p. 43) and pointed out the importance of including other measures of noticing (e.g., think-aloud, stimulated recalls).

In previous output and noticing studies, only two studies used stimulated recalls to investigate output-induced noticing (Uggen, 2012; Zalbidea, 2021). Uggen (2012) was a conceptual replication of Izumi and Bigelow (2000) using both underlining and stimulated recalls to investigate output-induced noticing. In addition to adding the retrospective verbal reports to the research design, she also examined how the complexity of the target linguistic forms (i.e., the English present and the past hypothetical conditional) could impact ESL learners' noticing behaviors in the subsequent input and the learning of each target form differently. Thus, the participants were randomly assigned to one of the three conditions: two output conditions

targeting either the present (EGpres, $n = 10$) or the past hypothetical conditional (EGpast, $n = 10$) and a control condition that still engaged in an output task but with no target form in focus (CG, $n = 10$). The learning gains were assessed through a written picture-cued production test and their performances on the essay writing task. The results revealed that the EGpast, which focused on the more morphologically complex form, demonstrated greater noticing in the subsequent input and learning gains compared to the CG. The output did not increase the EGpres's attentional allocation on the target form. Thus, the complexity of the target form was found to be an important moderating variable for output-induced noticing, suggesting that a morphologically more complex and difficult linguistic form (i.e., the past hypothetical conditional) may have turned out to be more salient for the learners and thereby triggered the EGpast group's noticing in the subsequent input. In contrast, the present hypothetical conditional may have been less salient for the participants of the EGpres group and thus directed their attention to semantic (lexical) elements in the text rather than the grammatical form. The study also highlighted the limitation of underlining as “a relatively uninformative source of evidence for studying noticing” (Uggen, 2012, p. 524).

Having reviewed previous output and noticing studies, the most recent and comprehensive study was Zalbidea (2021), which investigated the mechanisms of the noticing-triggering function of output and the roles of output in the processes of L2 acquisition in relation to two potential moderating variables: the perceptual and functional saliency of the target linguistic form (i.e., the Spanish simple future tense and the indirect object clitic) and the modality of L2 output (i.e., oral or written output) while addressing multiple methodological limitations identified in previous output and noticing studies. Furthermore, this study also investigated the levels of linguistic analyses that learners engaged in (i.e., depth of processing)

during the subsequent input processing depending on the treatment conditions. In this study, 88 beginner-level Spanish learners were divided into three instructional groups (i.e., Oral Output [$n = 28$], Written Output [$n = 30$], and Non-Output Groups [$n = 30$]) and engaged in two event-selection tasks, in which they were asked to select a logical follow-up event out of two possible options based on a brief prompt. During the treatment task, both output groups selected the follow-up event by orally describing the follow-up event or typing the description of the selected event depending on their output condition while the non-output group was asked to press a computer key to select their option instead of producing the target forms. After selecting the follow-up event, all the learners received feedback as a form of subsequent input, which provided them with an accurate targetlike model. Thus, the output conditions in this study consisted of the output-input cycle while the non-output condition consisted of input-input cycle. Their learning gains were assessed through an event-selection production test, which elicited the participants' target language production in both modalities and aural and written acceptability judgment tests. Five participants from each group ($n = 5$ for Non-Output, $n = 5$ for Oral-Output, and $n = 5$ for Written Output Groups) were asked to engage in retrospective stimulated recall sessions to examine the tendencies of each group's noticing behaviors during the respective task cycles (i.e., the output-input and the input-input cycles). The overall results showed supporting evidence of the noticing-triggering function of output and facilitative effects of L2 output on the learning of both salient and less-salient grammatical forms, which was evidenced by the deeper levels of linguistic analyses and the greater learning gains attained by both oral and written output groups. Comparing these two output modalities, the written output group demonstrated more sustained learning gains especially on the less-salient, challenging form (i.e., the indirect object clitic) on the results of the developmental tests also with greater incorporation rates of the

forms in their production during the treatment task than did the oral output group. As for the results of the noticing, which was examined through stimulated recalls, the two output modalities did not show substantial differences between the groups, both of which attained the highest level of noticing (i.e., successful integration) whereas more than half of the participants in the non-output group did not engage even in the lowest level of linguistic analyses (i.e., noticing and/or searching).

The author attributed the advantageous effects of the written modality to more processing time that was allowed by “the slower pace of writing and the nontransient visual quality” of the written modality (Zalbidea, 2021, p. 77). This increased time-affordance of the written modality may have enabled the learners to engage in grammatical encoding and monitoring, eventually increasing the learners’ sensitivity even toward less salient linguistic forms in the subsequent input (see further discussions on the potential roles of the written modality of output). As Zalbidea (2021) attributed the greater learning gains demonstrated by the Written Output group than the oral output group to the modality difference of output, the modality difference may play an important role in facilitating L2 grammar learning and can be an important moderating variable for L2 grammar acquisition through output.

Based on these reviews of the previous output and noticing studies, it was possible that learners’ output-induced noticing was not fully or accurately measured in previous output and noticing studies, most of which primarily relied on less sensitive measures of learner noticing, such as learners’ underlining, note-taking, and retrospective questionnaires. Even though Uggen (2012) and Zalbidea (2021) employed stimulated recalls, the verbal report still has its limitations as the measure of noticing. These measuring issues are discussed in the following section.

Measuring Issues of Learner Noticing

Considering the complex nature of output-induced noticing, which is a non-observable, internal cognitive phenomenon, it is valuable to empirically investigate whether and how the act of producing output induces learner noticing and promotes L2 learning by employing multiple measures that gauge learner noticing (Hanaoka & Izumi, 2021).

To summarize the measures used to capture the locus of learners' attention and the level of their awareness in previous studies (not only limited to output and noticing study here), the following noticing measures have been used: (1) diary entry (Schmidt & Frota, 1986); (2) underlining (e.g., Ghari & Moizadeh, 2011; Izumi & Bigelow, 2000; Russell, 2014; Shin, 2011; Song & Suh, 2008; Uggen, 2012); (3) note-taking (e.g., Hanaoka, 2007; Hanaoka & Izumi, 2012; Izumi, 2002; Kang, 2010; Leeser, 2008); (4) retrospective questionnaire (e.g., Izumi & Izumi, 2004); (5) think-aloud protocols (Hama & Leow, 2010); (6) stimulated-recall protocols (Godfroid et al., 2010; Uggen, 2010; Zalbidea, 2021); (7) a relatively new technique in this field, eye-tracking (Godfroid et al., 2010; Godfroid et al., 2013; Li & He, 2017; Godfroid & Uggen, 2013; Winke, 2013); and (8) finger-tracking (Godfroid & Spino, 2016).

These measures can be categorized into several types based on the characteristics and the strengths and limitations of these measures. As presented in Table 2.1, online measures provide information on what the learners are actually doing while they are engaging in the treatment task, thereby providing “relatively more substantial evidence of processing or processes” with higher internal validity (Leow, 2015, p. 137). In contrast, offline measures are post-treatment types of measures, which allow researchers to retrospectively examine the locus of learners' attention and awareness in detail. However, one of the biggest limitations of these offline measures is the inherent susceptibility to the participants' memory decay. The sensitivity of measures can also be

divided into either direct or indirect depending on how much the measure allows direct access to learners' ongoing internal processes (Izumi & Bigelow, 2000).

Table 2.1. *Noticing measures used in previous studies*

	<i>Sensitivity</i>	<i>Online</i>	<i>Offline</i>
Verbal	Direct	--	--
	Indirect	- Think-aloud - Note-taking	- Diary entry - Retrospective Questionnaire - Stimulated recall - Interview
Not verbal	Direct	- Eye-tracking	--
		- Finger-tracking	--
	Indirect	- Circling	--
		- Underlining	--

In addition to online/offline and direct/indirect classifications, noticing measures can also be classified based on whether the measure involves participants' verbal reports or not. Verbal reports allow researchers to closely explore learners' thought processes, which are otherwise unavailable (Gass & Mackey, 2017). However, veridicality (i.e., whether the verbal report is accurately reflecting the participant's thought processes that are being explored) and reactivity (i.e., whether the verbalization itself can influence the very thought processes under investigation) are the two major potential issues that always need to be considered regarding the use of verbal reports in SLA research. Even though online (concurrent) verbal measures are less susceptible to memory loss, they still have some limitations and difficulties in measuring what learners fail to report.

As reviewed above, Uggen (2012) and Zalbidea (2021) employed verbal reports through stimulated recalls, more detailed processes of learner noticing were highlighted in depth by the use of relatively sensitive verbal reports compared to previously employed less sensitive measures of learner noticing (e.g., underlining, note-taking, and retrospective questionnaires). Particularly, Zalbidea (2021) shed light on the depth of processing for output-induced noticing in

subsequent input. Although stimulated recalls provide more detailed accounts of learners' noticing behaviors compared to other less sensitive measures, the same weakness can be still pointed out for stimulated recalls, which is that what learners do not or fail to report or verbalize could not be measured regardless of its strengths in terms of the sensitivity and the depth of descriptive details. Furthermore, the biggest limitation of stimulated recalls may be the relatively small sample sizes due to the limited amount of data that can be dealt with by the researcher. For example, Uggen's (2012) study focused on ten participants for each of the three groups ($N = 30$) and Zalbidea (2021) conducted stimulated recalls with five participants for each group ($N = 30$). Therefore, it was still difficult to reach a generalizable conclusion based on the results of stimulated recalls due to the small sample sizes. Based on the advantages and disadvantages of these measures used in previous noticing studies in SLA, Godfroid et al. (2013) claimed that while learners' awareness is better measured through learners' verbal reports, such as stimulated recall and think-aloud protocols, the amount of attention and the locus of attentional resource needs to be measured through more sensitive online objective measures (e.g., eye-tracking).

Eye-tracking is a direct, online measure that can objectively capture L2 learners' eye movement on a computer screen. The cognitive rationale of the eye-tracking measure is the eye-mind link, which means that "eye movement can offer a window into the cognitive processes and knowledge that participants use to accomplish a particular task or goal" (Godfroid, 2019, p. 44). With this rationale, eye-tracking has been used to explore the relationship between learners' noticing and L2 learning both for vocabulary and grammar by examining the place and the amount of learners' attentional allocation while they are engaging in learning tasks (e.g., Godfroid et al., 2010; Godfroid et al., 2013; Godfroid & Uggen, 2013; Indrarathne & Kormos, 2017; Issa & Morgan-Short, 2019; Jung & Revesz, 2018; Li & He, 2017; Winke, 2013).

As for the eye-tracking measure, however, no published research attempted to use the eye-tracking measure at this point to test the noticing function of output regardless of its methodological strengths. The only study that investigated the noticing function of output, to my knowledge, was Li and He (2017), which partially replicated Izumi and Bigelow (2000). This study compared the amount of noticing induced by written output through a picture-cued written production task between two conditions (Output Group [$n = 33$] and Non-Output Group [$n = 12$]). They reported a significant increase in the output group's noticing, which was operationalized by the fixation counts and the total fixation duration, after producing output (i.e., in the second essay reading). They also found substantial learning gains in the results of the pretest and posttest. However, the study was still in progress at the time and the researchers were collecting data for more participants for the non-output group. Thus, the number of participants was very limited at the time. Therefore, it is valuable to test the noticing-triggering function of output and further investigate the relationship between output and the learning of L2 grammar with a large number of participants using an eye-tracking measure. Since eye-tracking measures can provide fine-grained, online objective data on L2 learners' attentional allocations, eye-tracking data on L2 learners' output-induced noticing may clarify the detailed mechanisms of the noticing-triggering function of output and further provide additional accounts for the inconclusive results of previous output and noticing studies due to their use of less-sensitive, indirect measures of noticing, such as underlining (e.g., Ghari & Moizadeh, 2011; Izumi & Bigelow, 2000; Russell, 2014; Shin, 2011; Song & Suh, 2008; Uggen, 2012), note-taking (e.g., Izumi, 2002; Kang, 2010; Leeser, 2008), and retrospective questionnaire (e.g., Izumi & Izumi, 2004).

Potential Roles of Output Modalities on Learner Noticing and Grammar Learning

Regarding the modality difference of L2 output, beneficial roles and the potentials of the written modality in L2 grammar learning started to gather considerable attention among ISLA/writing researchers particularly in recent years (e.g., Cumming, 1990; Gilabert et al., 2016; Harklau, 2002; Manchón, 2014; Manchón, 2011; Polio, 2020, 2022; Vasylets & Gilabert, 2022; Williams, 2012; Zalbidea, 2020; Zalbidea & Sanz, 2020). Although previous reviews discussed the potential benefits and roles of the written modality in L2 grammar learning from the *writing to learn* language perspective (e.g., Cumming, 1990; Gilabert et al., 2016; Harklau, 2002; Manchón, 2014; Polio, 2020, 2022; Williams, 2012), empirical evidence of the unique facilitative roles of the written output modality compared to the oral modality is still very limited particularly in relation to output-induced noticing (see Zalbidea, 2020, 2021). Therefore, questions regarding the effects of output modality, whether written or oral, on the development of certain linguistic forms have great theoretical importance and thereby need to be further examined empirically. Pedagogically, examining the impact of output modality difference is also valuable in that the findings and implications from studies that investigate potential effects of output modalities on learner noticing and grammar learning enable L2 teachers and curriculum designers to make theoretically- and empirically valid pedagogical decisions (e.g., whether an output activity should be conducted in which modality for what purposes for students with varying proficiency levels within limited amount of class time and schedules) and also enable teachers to systematically evaluate and revise their own/others' output-based teaching practices.

Regarding the potential roles of the written modality of output in the processes of the overall L2 acquisition, Williams (2012) reviewed and discussed in comparison to the oral output modality. In her comprehensive reviews of written output, two features of written modality (i.e.,

the slower pace of the delivery during the production and the permanence of the record of the written output) were identified as inherent beneficial features that differentiate the written modality from the oral modality. According to Williams (2012), “these two features permit more learner control over attentional resources as well as more need and opportunity to attend to language both during and after production” (p. 322) (see Figure 2.4). As described in Figure 2.4, both of these inherent features of the written output decrease the cognitive load on learners’ working memory (WM) while producing output in the written modality and thus provide more opportunity to notice holes and gaps in their current linguistic ability (Doughty, 2001; also see Izumi, 2013 for the four different types of noticing induced during and after producing L2 output). While learners are engaging in these noticing behaviors in the written modality, they have more time for processing and higher demand for linguistic accuracy, both of which are likely to promote their cognitive processes for planning and monitoring by consulting and reflecting on their explicit knowledge to focus on their problematic forms. Thus, learner noticing on these problematic linguistic forms is more likely to be promoted in the written modality than the oral modality of output. In contrast, Gilabert et al. (2016) pointed out the limitations of the oral modality of output by claiming; “Because of the evanescent nature of oral output, learners may register linguistic inconsistencies only transiently, with the results that even if noticed, the noticed elements can fade away from the speaker’s WM without any further processing” (p. 127).

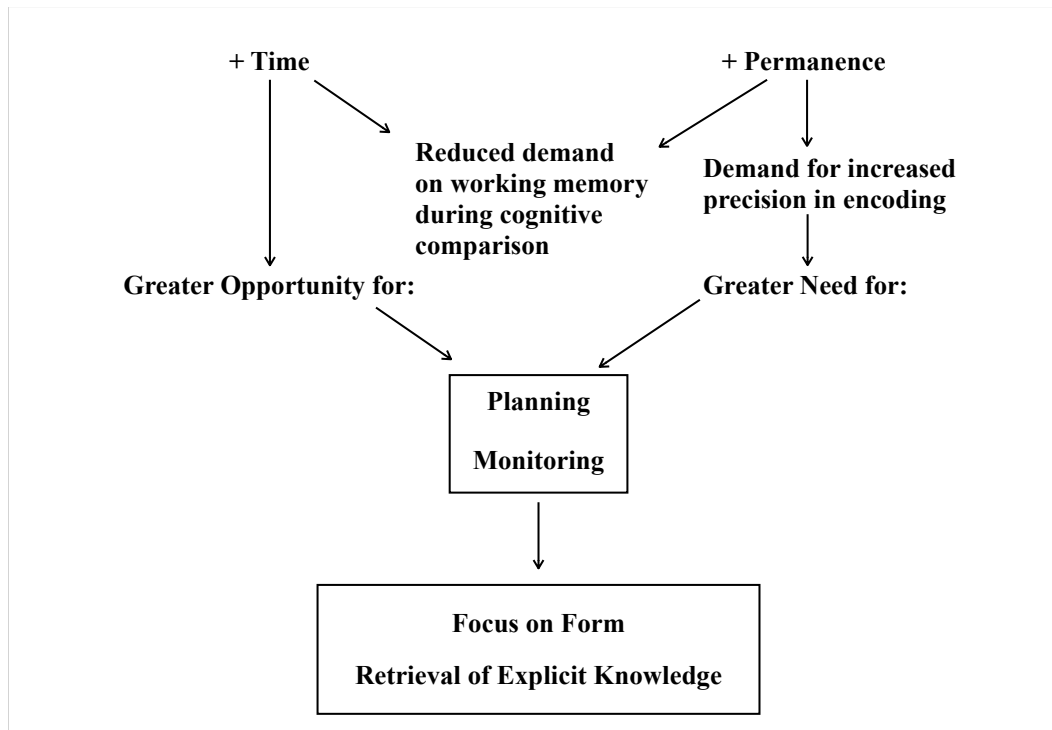


Figure 2.4. Williams' inherent features of written production and their effects (Adopted from Williams, 2012, p. 323)

Having reviewed the potential roles of the modality differences of L2 output from the perspective of cognitive processes, previous studies also pointed out differential effects of the output modality from learners' psychological perspectives. Regarding L2 learners' perceptions about the modality difference of output, previous studies reported higher task task-induced stress and anxiety for the oral output modality than the written modality (e.g., Baralt, 2013; Cho, 2018; Zalbidea, 2020). For example, Zalbidea (2020) conducted a post-task questionnaire to investigate L2 learners' perceptions about output-task demands. Questionnaire ratings of the two groups of learners (i.e., Oral Output [$n = 26$] and Written Output Groups [$n = 28$]) were compared in terms of various output-task features (e.g., perceived mental effort, task difficulty, stress levels, perceived task performance, timing/rushedness, task interest, anxiety, perceived linguistic difficulty in terms of output demands, etc.). Although no significant group differences were indicated on any of these output-task features, the oral output group indicated higher ratings

particularly for the ratings of task-induced stress levels and anxiety, $d = .44$ and $d = .33$ respectively, compared to the ones of the written output group.

Based on these reviews and theoretical postulations about the effects of output modality difference on learner noticing and L2 grammar learning, it seems that written output may be more facilitative than oral output in promoting learner noticing on and acquisition of problematic linguistic forms. Regarding the impacts of output modality difference (oral or written output) on L2 grammar learning, however, previous empirical studies primarily focused on the written modality of output (e.g., Alsulami, 2016; Basterrechea et al., 2014; Ghari & Moinzadeh, 2011; Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999; Kang, 2010; Leeser, 2008; Li et al, 2016; Li & He, 2017; Muraoka, 2006; Russell, 2014; Shin, 2011; Shintani, 2019; Song & Suh, 2008; Uggen, 2012). Only a few studies investigated the impact of the oral output modality on grammar learning (e.g., Izumi & Izumi, 2004) and compared its effects with the ones of the written modality (e.g., Muranoi, 2007b; Zalbidea, 2020, 2021).

As indicated above, Izumi and Izumi (2004) reported higher output-induced noticing for the oral output group but failed to show beneficial roles of oral output on L2 grammar learning. Due to the methodological issue of their output task design, they were unable to specify why the oral output itself had limited effects in relation to the roles of output modalities. Hence, it was unclear which modality was more or less beneficial in promoting learner noticing and grammar learning based on the psycholinguistic processes of grammar learning. Another study that compared the effects of output modalities on L2 grammar learning was Muranoi (2007b), which was a product-oriented study that investigated the effects of text-reconstruction task termed *guided summarizing* (GS) on the acquisition of English perfect passive but without examining learner noticing while learning was taking place. In this study, 40 Japanese EFL university

students were assigned to either the written plus oral GS or the written-only GS groups. During the treatment sessions, learners were asked to summarize a short reading text using a concept map (i.e., a semantic representation of words that indirectly guides learners to use a specific target form). For each reading passage (in each session), the learners were directed to work on the GS task twice. Both groups engaged in the first GS trial in the written mode and then performed the second trial in a different respective modality based on the instructional condition (either oral or written GS). Between the first and second GS performances, an interval reflection time was provided to have learners reflect on their first GS performance with access to the original reading text, which aimed to facilitate output-induced noticing and cognitive comparisons. The results of the oral and written sentence completion tests indicated significant posttest gains for both groups but with higher effect sizes for the written plus oral GS group (medium to large effect sizes) than the written-only GS group (small effect sizes) (see Muranoi, 2007b for all the effect sizes on both tests' gain scores). The results seemed to show higher learning gains if learners engaged in oral output in the second trial. However, Muranoi (2007b) did not provide any specific explanations on the potential effects of modality difference because the results did not show any significant differences between the groups. Also, the lack of process-oriented measures in this study (e.g., eye-tracking, stimulated recalls, or think-aloud protocols) made it difficult to interpret and specify the roles of output modality differences in the processes of grammar learning rather than solely relying on indirect theoretical speculations based on the results of the pretests and the posttests. Furthermore, although this study incorporated the output modality into the task design to compare the output modality difference, the higher learning gains were attained by the written plus oral output group as the results of engaging in both written and oral output. Due to the methodological issues of these composite

variables, a pure comparison of the modality difference was difficult and thus the roles of output modalities were still unclear.

As indicated above, Zalbidea (2021) was the first and only noticing and output study that showed advantageous effects of the written output modality in comparison to the oral modality on L2 grammar learning. This study clearly demonstrated greater and more sustained effects of written output modality than oral output especially on the less-salient, challenging form (i.e., the indirect object clitic). Concerning the noticing behaviors, however, both output modality groups indicated equally high levels of linguistic analyses of the target forms compared to the non-output group. In other words, both output modalities equally promoted deeper levels of learner noticing, supporting the noticing-triggering function of output. Again, the biggest limitation of her study was the limited number of participants who participated in the stimulated recall sessions (five participants from each condition: $n = 5$ for Non-Output, $n = 5$ for Oral-Output, and $n = 5$ for Written Output Groups). Therefore, it is valuable and is needed to closely examine the potential roles of output modalities in the processes of L2 acquisition (e.g., learner noticing and overall grammar learning attainments) through a fine-grained, online objective measure on L2 learners' attentional allocations with larger sample sizes so that the results can indicate generalizable tendencies of the modality effects of L2 output on learner noticing and grammar learning. Addressing the limitations of these previous studies would further advance the theoretical understanding of Swain's Output Hypothesis, the roles of L2 output, and the noticing-triggering function of output in relation to the overall processes of L2 acquisition.

The Present Study

As reviewed in this section, the roles of L2 output on learner noticing and grammar acquisition are still unclear due to the mixed results indicated by previous output and noticing

studies, methodological limitations of noticing measures, and varying operationalizations of output in relation to modality differences. Therefore, whether and how specific types of L2 output contribute to the overall processes of L2 acquisition has not been fully understood regardless of the commonly accepted, usefulness of output in L2 learning among SLA/ISLA researchers and L2 teachers. Based on the limitations of previous empirical studies, theoretical reviews, and their methodological recommendations regarding the use of more sensitive measures of learner noticing (e.g., Godfroid et al., 2013; Hanaoka & Izumi, 2021; Uggen, 2012; Zalbidea, 2021), the present study examined whether and to what extent producing output can induce learner noticing in the subsequent input and facilitate their acquisition of the target grammatical forms as the noticing-triggering function of Swain's Output Hypothesis proposed through employing a more sensitive online objective measure of learner noticing (i.e., eye-tracking) along with grammar development measures (i.e., an oral elicited imitation test [OEIT] and a written picture description test [WPDT]). Addressing these questions with a hybrid design of process- and product-oriented research is critical to looking into the black box of L2 learning and elucidating the roles of L2 output and learner noticing in the overall acquisition processes depicted in Gass' (1997) integrated model of L2 acquisition.

Research Questions and Hypotheses

To address the gaps discussed above, the present study investigated output-induced learner noticing in the subsequent input to answer the research questions listed below. For these research questions, corresponding hypotheses are provided based on the theoretical postulations and the empirical findings of previous output and noticing studies reviewed above.

RQ1: To what extent does producing output induce learners' noticing of the target linguistic form (i.e., English past counterfactual/hypothetical conditional) in the subsequent input?

RQ2: How does the modality of output affect the output-induced noticing of the target form in the subsequent input differently?

Based on Swain's Output Hypothesis and previous output/noticing studies, it can be hypothesized that both output groups would show greater noticing of the target linguistic form in the subsequent input compared to the input-only condition (i.e., Input-Only Group), which could be indicated by both early and late measures of noticing (i.e., eye-tracking).

If the first research question is confirmed, both output conditions would be compared based on the modality difference of output (either written or oral output). Based on the reviews and the findings from previous empirical studies that highlighted the beneficial roles of writing for L2 grammar learning (e.g., Cumming, 1990; Gilabert et al., 2016; Harklau, 2002; Manchón, 2014; Polio, 2020, 2022; Williams, 2012; Zalbidea, 2020, 2021), the more processing time and the permanent records of output producing provided by the written modality would enable the written output group to spend more cognitive resources to engage in deeper structural analysis of the target linguistic form with more cognitive effort than the oral output group. As a result of this deeper processing, the written output group may pay more attention to the target linguistic form with longer eye-fixation durations on both early and late measures in the subsequent input than the oral output group would do.

RQ3: To what extent does producing output contribute to the learning of the target linguistic form (i.e., English past counterfactual/hypothetical conditional)?

RQ4: How does the modality of output affect the learning of the target linguistic form?

RQ5: Is the amount of output-induced noticing associated with the overall learning of the target linguistic form?

For Research Questions 3 and 4, similar hypotheses were postulated. Both output groups may demonstrate greater learning gains both on the written picture description test (WPDT) and the oral elicited imitation test (OEIT) than the input-only group. If the written output group indicates a greater noticing of the target linguistic form, it can be hypothesized that the written modality can facilitate the greater learning of the target linguistic form than does the oral output group. The input-only group would be assumed to show negligible learning gains on both grammar developmental measures (i.e., WPDT and OEIT).

To answer Research Question 5, the association between learner noticing in the subsequent input measured through eye-tracking and the development of target linguistic knowledge measured through the WPDT and the OEIT was examined through correlation analysis and multiple regressions. The hypothesis for Research Question 5 was postulated that the learning gains on both developmental tests would be associated with the amount of attention directed to the target form (i.e., both the early and late measures) while processing the subsequent input.

CHAPTER 3: METHOD

The present study investigated whether and to what extent producing L2 output could induce learner noticing and facilitate the learning of the target grammatical form (i.e., English past counterfactual/hypothetical conditional) by employing a direct online objective measure of noticing (i.e., eye-tracking) along with two developmental measures of L2 grammar knowledge (i.e., an oral elicited imitation test [OEIT] and a written picture description test [WPDT]). Figure 3.1 presents the overall experimental design of the present study. This section provides the following methodological details: participants, target linguistic form, experimental procedures, target tasks, measuring instruments, coding and scoring procedures, and data analysis.

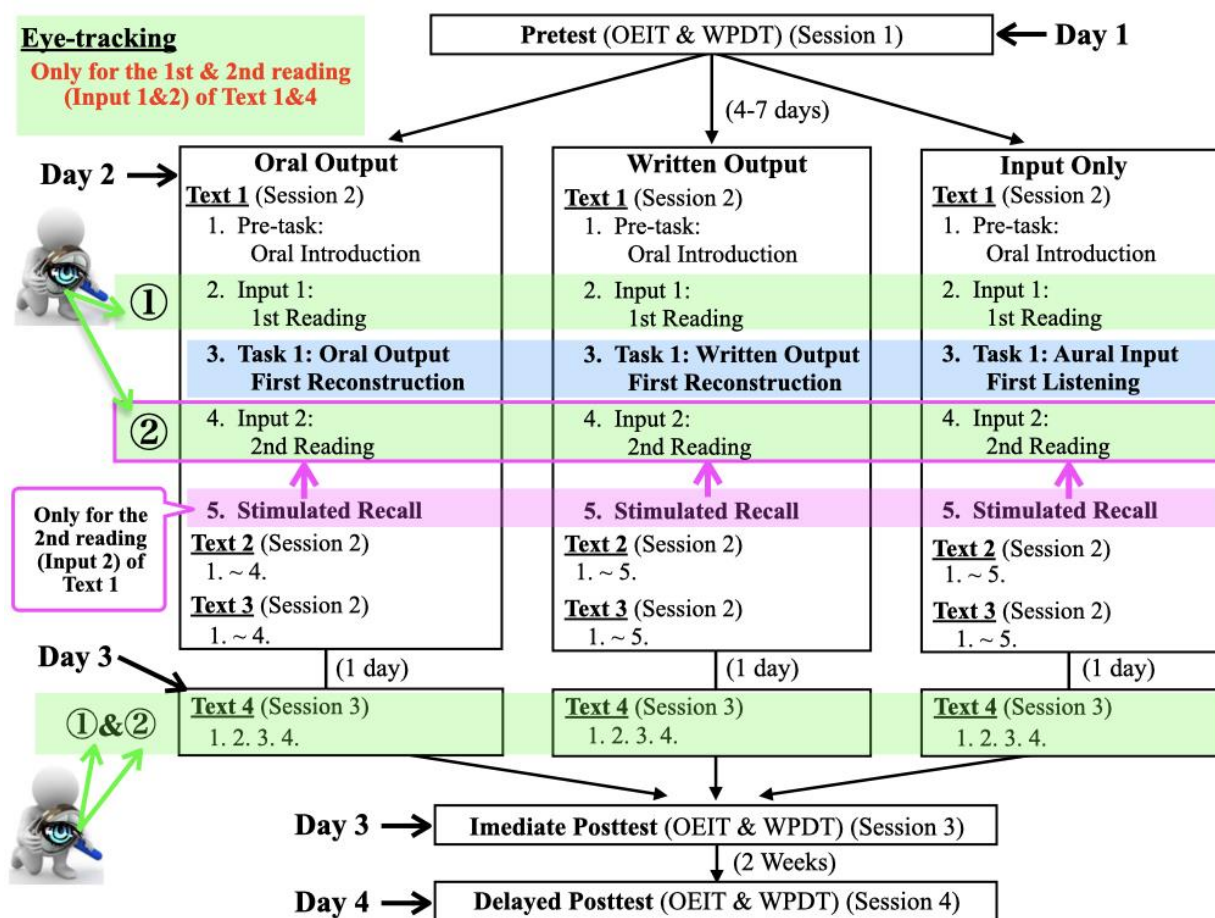


Figure 3.1. The overall research designs

Participants

The participants of this study were 117 international undergraduate and graduate students who were studying various disciplines at Michigan State University. All the participants' English proficiency levels were classified as B2 level in the Common European Framework of References for Languages (CEFL) based on their English proficiency test scores. To recruit participants whose English proficiency level was classified as the B2 level in the CEFL, the following English proficiency tests were used to convert the participants' English proficiency test scores to the classifications of the B2 level in the CEFL: TOEFL iBT (75-95) (Papageorgiou et al., 2015), IELTS (5.5-6.5) (Hawkey & Barker, 2004; Lim et al., 2013), TOEIC (745-945) (Schmidgall, 2021; Tannenbaum & Wylie, 2008), and the Duolingo English test (100-125) (Duolingo English Test, 2021). Any international students whose English proficiency level was indicated below or above these score ranges were not recruited in this study. However, it turned out that two participants were mistakenly included in the participants, and thus both of them were excluded from the analysis.

In addition to the converted English proficiency levels in the CEFL, the participants' vocabulary knowledge level was also examined through Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT), which was indicated as a reliable predictor of L2 learners' receptive language proficiency (Ha, 2021) (see Appendix A for the entire test). All the participants' averaged score on the UVLT (1K-5K frequency level) was 138.86 (92.57%, $SD = 12.49$, $Min = 110$, $Max = 150$) out of the maximum score of 150. As for the most frequent 3000-word families on the test (1K-3K frequency level), their average score was 90.22 (97.01%, $SD = 2.75$, $Min = 82$, $Max = 93$) out of the maximum score of 93, suggesting that the participants may not have had difficulty reaching a fair comprehension of the reading texts used during the

treatment sessions because all the reading texts consisted primarily of 1K-3K vocabulary (see Table 3.1. for the vocabulary level of the reading texts used in the treatment sessions). The vocabulary test results for each instructional group are summarized in Table 3.1. All the participants whose 1K-3K and 1K-5K scores exceeded more than 2SD were excluded from the final samples and seven participants were excluded based on these criteria.

To check the group equality in terms of their vocabulary knowledge (or their receptive proficiency), a one-way ANOVA was conducted both on their 1K-3K and 1K-5K scores on the UVLT, indicating no significant differences on 1K-3K ($F(2, 80) = 2.49, p = .09, \eta^2 = 0.06$) and 1K-5K ($F(2, 80) = 1.89, p = 0.16, \eta^2 = 0.05$).

Table 3.1. Results for Webb, et al.'s (2017) Updated Vocabulary Test for each group

<i>Group</i>	<i>Freq. Level</i>	<i>M (%)</i>	<i>Mdn %</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input	1K-3K	91.04 (97.89%)	98.92%	86	93	2.18	[90.20, 91.88]
	1K-5K	141.31 (94.21%)	95.33%	121	150	7.82	[138.31, 144.31]
O_Output	1K-3K	89.41 (96.14%)	95.70%	82	93	2.95	[88.33, 90.49]
	1K-5K	136.93 (91.29%)	92.00%	124	150	6.83	[134.44, 139.42]
W_Output	1K-3K	90.29 (97.08%)	97.85%	83	93	2.87	[89.23, 91.35]
	1K-5K	138.57 (92.38%)	94.00%	10	149	10.15	[134.81, 142.33]

Note. The possible highest score 1K-3K = 93, The highest score 1K-5K = 150, Freq. Level:

Vocabulary frequency level

The participants' nationality with all the groups combined included: China ($n = 16$), Taiwan ($n = 10$), Indonesia ($n = 9$), South Korea ($n = 8$), Japan ($n = 7$), Brazil ($n = 6$), Kazakhstan ($n = 4$), Bangladesh ($n = 3$), Malaysia ($n = 3$), Turkey ($n = 3$), Colombia ($n = 2$), Spain ($n = 2$), Thailand ($n = 2$), Vietnam ($n = 2$), Argentina ($n = 1$), Belarus ($n = 1$), Chile ($n = 1$), Costa Rica ($n = 1$), Czech Republic ($n = 1$), France ($n = 1$), Germany ($n = 1$), Honduras ($n = 1$), Iran ($n = 1$), Pakistan ($n = 1$), Panama ($n = 1$), Paraguay ($n = 1$), Peru ($n = 1$), Russia ($n = 1$), and Saudi Arabia ($n = 1$). All of them received their formal English education in junior high or

senior high schools in an English as a foreign language (EFL) context and those who had stayed in English-speaking countries or English as a second language (ESL) contexts (e.g., the US, Canada, Australia, etc.) for five years or more were not recruited as the participants of this study. The percentage of the length of their residence in English-speaking countries was the following: less than 1 year ($n = 54$, 58.06%), 1-2 years ($n = 19$, 20.43%), and 3-5 years ($n = 20$, 21.51%). Their majors included various subjects (e.g., Accounting, Biology, Economics, Kinesiology, Law, Statistics, Veterinary Medicine, etc.) but other than English teaching or TESOL (Teaching English to Speakers of Other Languages) (see the discussions in the Revisions Based on the Pilot Study section).

Once the participants were recruited with the initial screening, they were randomly assigned to one of the following three different instructional conditions: Oral Output Group ($n = 32$), Written Output Group ($n = 31$), and Input-Only Group ($n = 30$). As presented above, the equivalence of their vocabulary test performances and their pretest performances of both linguistic measures (i.e., the OEIT and the WPDT) were examined (see Table 3.1. for their UVLT scores, also see Tables 4.19. and 4.25. for their pretest scores). Since the primary aim of this study was to examine whether and to what extent producing L2 output induced learner noticing of the target linguistic form in the subsequent input and the overall learning gains, participants who had already fully acquired the target linguistic form at the point of pretest were removed from the analysis. The maximum cut-off point was set to 80 percent of the pretest scores on both the OEIT and the WPDT and those who exceeded the cut-off point on both tests were excluded from the analysis, removing three participants for the final analyses. Those who missed any of the four data collection sessions and those whose eye-movements were not successfully tracked by the eye-tracker ($n = 10$) were also removed from the participants included

in the final analyses, leaving 83 participants in the final analyses.

All participants had either normal or corrected-to-normal eyesight. The participants who completed all the data collection sessions received \$45 as monetary compensation. Before starting all the data collection, the research proposal and the experimental procedures were accepted by the Institutional Review Board (IRB) at Michigan State University (see Appendix B for the recruiting flyer used for the recruiting).

Target Form

The target linguistic form of this study was the English past counterfactual/hypothetical conditional (e.g., *If Steve Jobs had not learned calligraphy, the first Mac computer would not have had wonderful fonts with beautiful calligraphy.*). This linguistic form is generally challenging even for advanced ESL/EFL learners due to its syntactic and semantic complexities (Celce-Murcia & Larsen-Freeman, 2016, Izumi et al., 1999; Izumi & Bigelow, 2000; Shintani, 2019; Uggen, 2012). To accurately use this form, learners need to process multiple linguistic elements at the same time. In particular, learners need to process both the main clause and the subordinate clause to express the hypotheticality and the past-time reference using an additional marker in the form of past perfect.

Although the participants were recruited from groups of international undergraduate and graduate students who were assumed to possess relatively high levels of functional skills in English, it was still unlikely that the participants of this study had solid control over this structure even though all of them possessed functional English proficiency (i.e., the B2 level in CEFL) (see the results of the pretests on both OEIT and WPDT in Chapter 4). As shown in the results of the pretests (OEIT and WPDT), the participants still failed to accurately encode this target grammatical structure at the point of the pretest, showing about 40 percent mean accuracy on

both OEIT and WPDT, even though their test performances were calculated with the interlanguage (IL) scoring system, which added up partial scores for each of the seven elements of English past hypothetical conditional, rather than the strict target-like (TL) scoring system, which ignores the participants' partial and emerging knowledge of the target form unless all the seven features and components of the form are accurately used (see the scoring procedures for the details of the IL scoring system). Furthermore, Uggen's (2012) participants were also recruited from the same group of population at the same university and showed limited linguistic knowledge of this target form at the point of the pretest. Thus, the target linguistic form was considered appropriate to examine output-induced noticing in the subsequent input after engaging in the output task during the treatment sessions.

Procedures

The present study was conducted in a pre-posttest design along with during-task processing measures (i.e., eye-tracking and stimulated recalls). The data collection of this study followed the procedures presented in Table 3.2.

At the beginning of the first session (Session 1), all the participants were asked to engage in a practice passage reading task using the eye-tracker for the purpose of getting the participants familiarized with reading a passage while their eye-movements were being tracked by the eye-tracker and also for the purpose of checking whether the participant's eye movements could be tracked by the eye-tracker. If the participants' eye movements were not accurately tracked at this practice stage due to failed calibrations and validations or any other unknown technical difficulties, which were likely the case for those who were wearing thick glasses or strongly corrected contact lenses, the rest of the entire data collection sessions were canceled for that type of participants. For these reasons, nine students' sessions were canceled at the stage of this

calibration checking (or the practice reading). The passage used for the practice reading task did not contain any instances of the target grammatical exemplars (see Appendix C for the passage used in the practice reading). After completing the initial practice reading, all the participants took the pretests (both the OEIT and the WPDT) and Webb et al.'s (2017) Updated Vocabulary Levels Test (UVLT) and completed a background questionnaire using a computer in the eye-tracking laboratory.

Table 3.2. *Experimental Procedures*

Week 1 (Day 1)	Session 1: (in the lab)	Practice reading with eye-tracking Pretest (OEIT, WPDT, Vocabulary test, and Background questionnaire)
Week 2-1 (Day 2)	Session 2: (in the lab)	Treatment 1 Text 1 1. Pre-task (Oral introduction) 2. Input 1 (First reading) ← Eye-tracking 3. Task (Oral output, Written output, or Aural input) 4. Input 2 (Second reading) ← Eye-tracking 5. Stimulated recalls with Input 2 eye-tracking data Text 2 (Tasks, Stimulated recall, Eye-tracking) Step 1-4 ← No eye-tracking
Week 2-2 (Day 3)	Session 3: (in the lab)	Treatment 2 Text 3 Step 1-4 ← No eye-tracking Text 4 1. Pre-task (Oral introduction) 2. Input 1 (First reading) ← Eye-tracking 3. Task (Oral output, Written output, or Aural input) 4. Input 2 (Second reading) ← Eye-tracking Immediate posttests (OEIT and WPDT)
Week 5 (Day 4)	Session 4: (online)	2-week delayed posttest (OEIT and WPDT)

Next week, they participated in the first and the second treatment sessions in the eye-tracking laboratory (Session 2 and 3). At the end of the second treatment session (Session 3), they also engaged in the immediate posttests. Two weeks after the immediate posttests, the

delayed posttests were conducted (Session 4) online using Zoom because they only needed to engage in the OEIT and the WPDT in this session using their computer. While each participant was working on both the OEIT and the WPDT in the final session, their computer screens were shared with the present researcher to make sure that each of them was engaging in the task as they did in the previous testing sessions. Hence, the participants were required to come to the eye-tracking laboratory three times in total for the first three sessions (Sessions 1, 2, and 3) and took the delayed posttests online via Zoom for the final session (Session 4).

Treatment Tasks

The participants were assigned to one of the three instructional conditions: Oral Output Group, Written Output Group, and Input-Only Group. As shown in Figure 3.1, all the participants (1) listened to a pre-task oral introduction that provided background knowledge of the reading text that they read in the following treatment task; (2) read a short text while their eye-movements were tracked by the eye-tracker (Input 1); (3) engaged in each different treatment task as the first trial (i.e., oral output, written output, or input) (Task); (4) read the same text again while their eye-movements were tracked by the eye-tracker (i.e., subsequent input) (Input 2); (5) engaged in the stimulated recall using the recording of their eye-movements during the second reading (i.e., Input 2). Both output groups engaged in the same text-reconstruction task, but the modality of the reconstruction (i.e., output) was different either in an oral or written mode. Since four reading texts were introduced during the two treatment sessions (two texts for each treatment session [Session 2 and 3]), the participants followed the same four steps (Steps 1-4) described above for each reading text (Text 1-4). However, the eye-tracking was conducted only for the first text and the final text (Text 4), and the stimulated recall session was conducted only after Step 4 of the first reading text (Text 1) (see Table 3.3). Even though stimulated recalls

were conducted after completing all the tasks for Text 1, the results of these were not included in the current paper.

Pre-task: Oral Introduction

The first phase of the instructional treatment was an oral introduction, in which the participants received background information about the reading passages with some pictures of the person featured in each reading passage (see Figure 3.2 for example slides used during the oral introduction; also see Appendix D for the slides used in the oral introduction for all the reading texts; see Appendix E for the narration scripts used for the oral introduction for all the texts). The primary aim of the oral introduction was to provide the participants with necessary background information about the characters described in the reading texts so that the participants could fully understand the facts about the characters in the reading texts, which provided the basis for the hypothetical stories described in the reading texts.

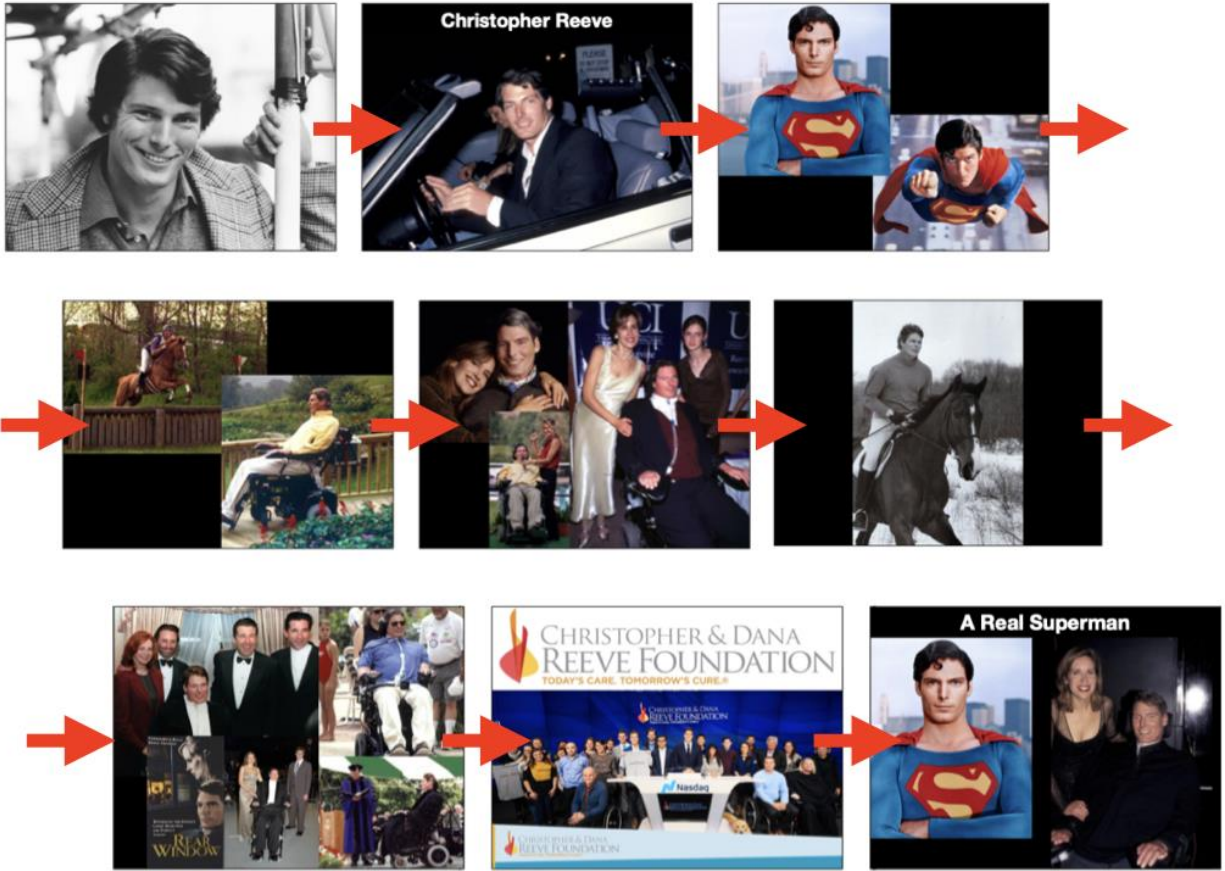


Figure 3.2. Example presentation slides used in the oral introduction for Text 4

Input 1 and 2: Text-Reading

The participants were asked to read a short-written text (i.e., input) that featured a famous person’s life history (i.e., Steve Jobs 1 and 2, Ichiro Suzuki, and Christopher Reeve) (see Table 3.3 and Appendix F). The reading passages were written by the researcher or were adopted from the materials of Izumi et al. (1999) after making some minor revisions considering the level of participants’ proficiency and the level of their vocabulary knowledge to maximize the participants’ comprehension.

Each text consisted of 110-118 words and contained four exemplar sentences that contained the target form (i.e., the past hypothetical/counterfactual conditional) (see Table 3.3 for the summary of the content and the number of the target form included in each text during the

instructional treatment sessions). The length of the passage was designed to be short enough to avoid overtaxing learners' attentional resources to reconstruct the passage but was still long enough to prevent mere verbatim memorization of the text without comprehending the meaning (or the message) of the text. For example, Izumi and Izumi (2004) asked their participants to reconstruct each sentence of a short passage sentence by sentence during the text-reconstruction task to reduce the processing load for the participants. However, their efforts to reduce the processing load ended up making the reconstruction task too easy for the participants, allowing simple, mechanical repetition without comprehending the meaning of the text (Izumi & Izumi, 2004). Therefore, the length of each reading passage of this study was determined based on the recommendations provided in the previous output studies that used text-reconstruction tasks and based on several iterations of piloting with international students from similar populations. The following is an example passage:

An example reading text (Text 4)

In 1995, Christopher Reeve fell off his horse. The accident left him paralyzed. If the horse had jumped over the hurdle successfully, Reeve would not have fallen off the horse. If his hands had been free, he would have landed safely. Despite the accident, he did not give up his hope to return to creative work and founded a charitable organization for spinal injury research. He would have given up all hope to live if his wife had not encouraged him to be strong. If he had felt discouraged, he would not have recognized his ability to raise money for medical research. Now, he is remembered by his fans as a real-life superman. (110 words)

Each passage consisted primarily of 1K-3K level vocabulary to avoid comprehension difficulties due to the lexical items that the participants had not acquired. Previous studies have reported, L2 learners tend to direct their attentional focus on lexical elements rather than grammatical elements while producing output (e.g., Hanaoka, 2006a, 2006b, 2007; Hanaoka & Izumi, 2012; Mackey, Gass, & McDonough, 2000; Swain & Lapkin, 1995; Uggen, 2012; Williams, 1999). Thus, it was crucial to use passages that consisted primarily of words that the learners were already familiar with (see Table 3.3).

Table 3.3. *Summary of the content and the number of target forms included in each text*

	1st Session		2nd Session	
	Text 1	Text 2	Text 3	Text 4
Topic	Steve Jobs	Steve Jobs	Ichiro Suzuki	Christopher Reeve
Word count	115 words	117 words	118 words	110 words
Vocab level (1-3K)	96.50%	96.60%	96.70%	95.50%
Target sentences	4 (57.14%)	4 (66.66%)	4 (57.14%)	4 (57.14%)
If clause first	2	2	2	2
If clause second	2	2	2	2

Task: Reconstruction Task (Oral and Written Output Groups)

After the participants completed the first reading (Input 1), the reading passage was collected by the researcher so that the text was not available to them while they were engaging in the text-reconstruction task. In this task, both output groups (i.e., Oral Output Group and Written Output Group) were asked to reconstruct the text as accurately as possible, with the help of descriptive pictures (see Figure 3.3). For each sentence that contained the instance of the past hypothetical conditional, a set of two pictures that depicted the sentence was provided (see Figure 3.3 for an example of the elicitation pictures) because it was possible that the participants could skip reconstructing the specific sentences that contained the target linguistic form if the picture cues were not used. Furthermore, without using these picture cues, it could be assumed

that the reconstruction task became cognitively too taxing for the participants' memory load. This issue was discussed by Izumi et al. (1999), in which the authors claimed that the text-reconstruction tasks without any memory support turned out to be too demanding for their participants' memory load and prevented them from engaging in careful analysis of the target linguistic form. For these reasons, the current study used picture cues to elicit the participants' reconstruction of the specific sentences that contained the target form.

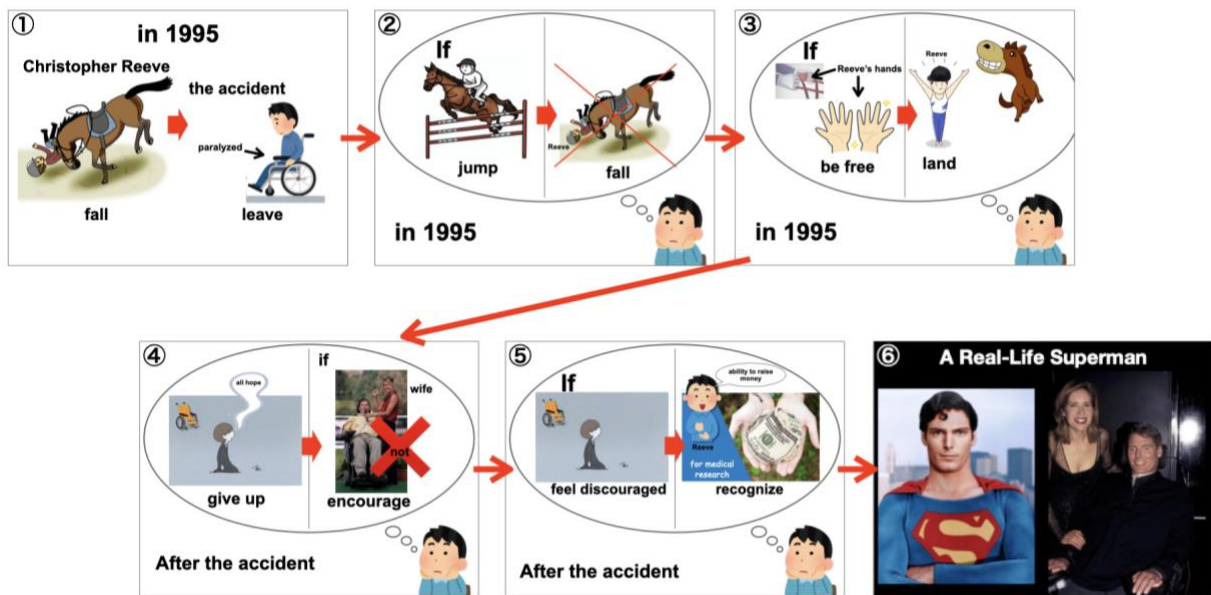


Figure 3.3. Example picture cues for the text reconstruction

The oral output group orally reconstructed the text, and the written output group typed their reconstruction on the computer screen. Their reconstruction performances were also checked and scored to compare their reconstruction performances between the oral and written output groups using the same scoring system presented in the Scoring Procedure section (also see Table 3.4). Comparing these two different modalities, reconstructing the passage in the written modality takes more time compared to the oral modality. However, each time on task was not

equalized in this study as Williams (2012) suggested, “I suggest that the longer time required for written production simply be accepted as an artifact of the modality” (p. 322).

Task: Listening to a Model Reconstruction (Input Group)

During the task phases (i.e., aural input), the input-only group listened to a model reconstruction read by an English native speaker. Since the output groups had opportunities to process the target structure four times during their reconstruction, the input group was also provided with the same number of opportunities to process the target linguistic form during the task phase, but the modality was in aural input.

Another reason to use the aural input for the input-only group during the task phase was to provide the learners with a different task from the one of the input phases but still with the same modality (the input-processing mode) because the task for the output groups during the task phase was different from the one of the input phases (i.e., Input 1 and 2). To minimize the differences between the output and the input conditions during the task phases, the input group did not work on the same reading task three times, but rather they listened to a model reconstruction performance sticking with the same input mode. The same descriptive pictures that the output groups used to reconstruct the text were provided as a form of picture-cued videos along with the model reconstruction narrations. Therefore, the total number of processing of the target exemplars during the instructional sessions are the same across the three instructional groups.

Measures

L2 Processing (Noticing) Measures

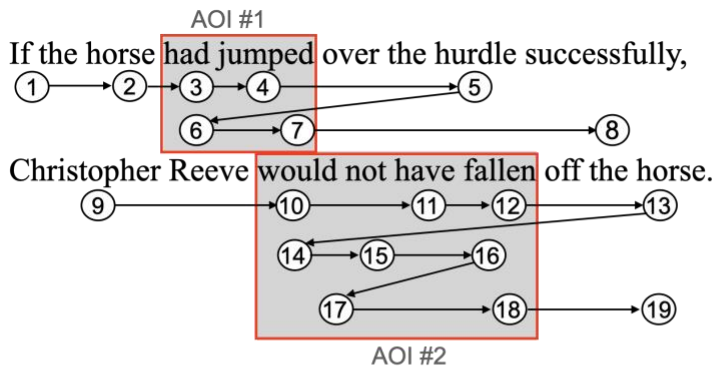
Eye-Tracking

All the participants' reading behaviors during the first and second reading (i.e., Input 1 and 2) of Text 1 and Text 4 were recorded with an eye-tracker, the Eye-Link 1000, during the treatment sessions (see Table 3.2 and Figure 3.1 for the timing of eye-tracking during the instructional treatments). In each text, there were four exemplar sentences, each of which consisted of one *if-clause* and the main clause. An area of interest (AOI) was created for each clause in one exemplar sentence (see Figure 3.4). Thus, in total, eight AOIs were included in each reading text (Text 1 and 4) (Appendix G for all the exemplar sentences and AOIs).

With the eye-tracking measure, the study aimed to examine (1) how long participants' eyes were fixated on the AOIs for the target linguistic form while they were reading the text before and after they worked on each respective task, which was operationalized as their noticing behaviors; (2) how their noticing behaviors differed depending on the treatment conditions while learning of the target grammatical form; and (3) whether or not these noticing behaviors were related to the overall acquisition of the target form. To answer these questions, the learners' reading behaviors and their eye-movements were recorded using both early and late measures of eye-fixation duration within each AOI, and their noticing behaviors were operationalized as first-pass reading time (FPRT) as the early measure and re-reading time (RRT) as the late measure (Godfroid, 2020).

The FPRT is one of the early measures that is defined as "the sum of all fixations recorded for a multi-word interest area up to the point when the eyes leave the interest area" (Godfroid, 2020, p.215). As shown in Figure 3.4, the FPRTs for AOI #1 and #2 are the sum of

the fixation durations of the fixations 3 and 4 and the fixations 10, 11, and 12, respectively. As one of the late measures, the RRT is defined as “the summed duration of all fixations in an interest area except for those fixations made during first pass” (Godfroid, 2020, p.215). Thus, the RRTs for AOI #1 and #2 are fixations 6 and 7 and fixations 14, 15, 16, 17, and 18, respectively (see Figure 3.4). These two early and late measures index different stages of sentence processing. The early measure represents the initial stage of sentence processing such as word recognition and lexical access, while the late measure indexes late stages of processing and is likely to signal an interruption to the normal reading process (e.g., effortful and nonautomatic processing) (Conklin et al., 2018; Godfroid, 2020; Maie & Godfroid, 2022). Therefore, it was crucial to use both fine-grained measures to closely examine how different task conditions (i.e., oral output, written output, and aural input) could influence the subsequent input processing and overall acquisition of the target linguistic form.



AOI #1

- First-pass reading time (Early) → ③ ④
- Re-reading time (Late) → ⑥ ⑦

AOI #2

- First-pass reading time (Early) → ⑩ ⑪ ⑫
- Re-reading time (Late) → ⑭ ⑮ ⑯ ⑰ ⑱

Figure 3.4. Example interest areas and fixations in one exemplar sentence

After completing the data collection, each participant's eye-tracking data was visualized and cleaned for the subsequent analysis using the EyeLink Data Viewer 3.2.1 computer software. In the process of data cleaning, first, the amount of track loss was inspected using a temporal graph. As for the fixations that were vertically drifted from the text lines, spotted drifted fixations were manually corrected (see Figure 3.5 for an example eye-fixation overlay image of the eye-tracking data). To deal with overly short and long fixations, fixations that were shorter than 50 milliseconds were merged or removed. On the other hand, fixations that were longer than 800 milliseconds were also removed because such long fixations are likely to represent participants' lapse of attention rather than their reading behaviors (e.g., Conklin et al., 2016; Godfroid, 2020). Therefore, these fixation duration thresholds were adopted for the data cleaning in this study (see Issa & Morgan-Short, 2019; Li & He, in press). After cleaning the data, total reading time and FPRT were extracted and calculated RRT based on these data. For thorough data exploration and checking, total fixation counts, first-pass fixation counts, and re-reading fixation counts were also extracted and calculated together with the focused variables presented above but these were not reported in this study.

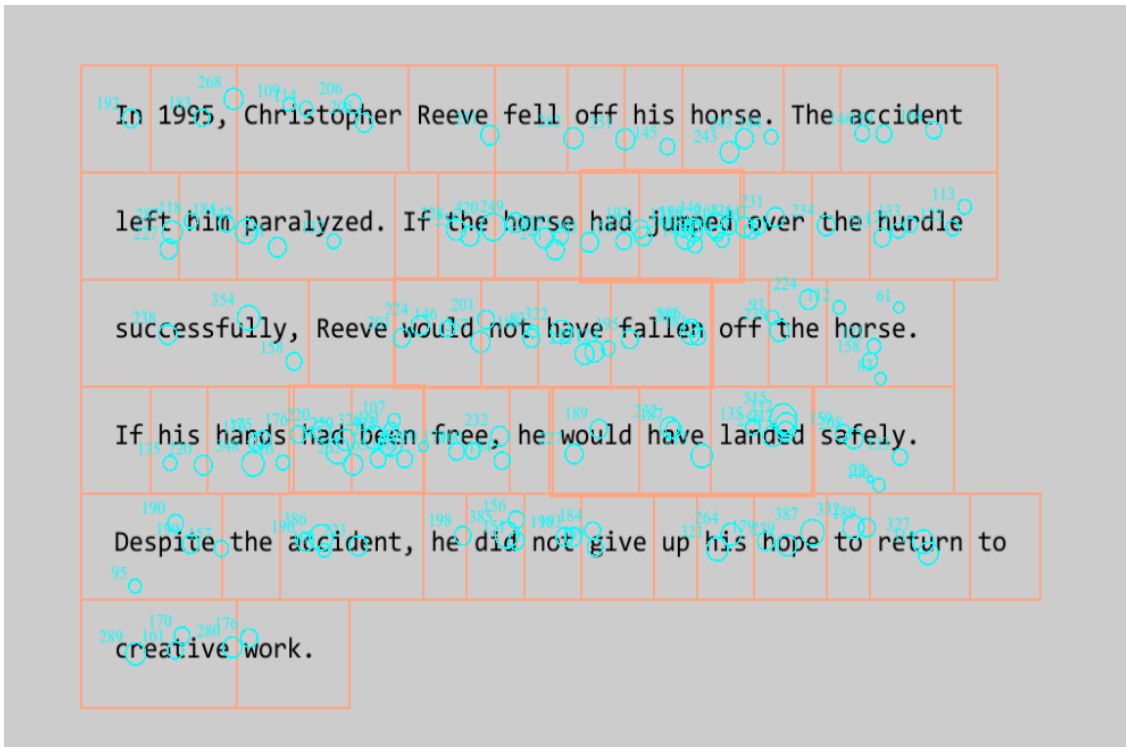


Figure 3.5. A sample eye-fixation overlay image of the first page (four sentences) of Text 4

L2 Knowledge Measures

Written Picture-Cued Description Test

To measure the effects of each treatment task on the development of the target grammatical knowledge, a written picture-cued description test (WPDT) was conducted. All the testing sessions were computerized and were implemented using an online survey software, *Qualtrics*. The test materials for the WPDT were adopted and revised from the picture-cued production test used in Izumi and Bigelow (2000) and Suh (2010) and also were newly created following these formats.

The WPDT consisted of one example section and three testing sections (Sections 1, 2, and 3). In each section, there was one passage that provided the background information for the pictures to be described at the beginning (see Appendix H). The following is an example introduction passage for one section of the WPDT:

An example introduction passage

Part 1 (Lisa)

Lisa had many options when she finished high school in 1992. However, she decided to work for an insurance company. The following pictures illustrate Lisa's other options and their potential consequences (results). She did not choose these options. In each pair, Picture A shows Lisa's option when she finished high school in 1992, and Picture B shows the potential consequence (result) of the option.

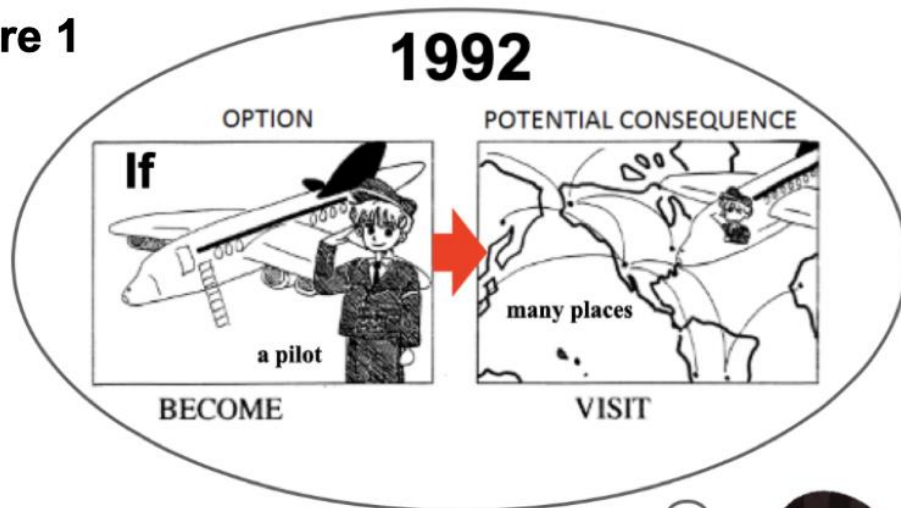
In each section, there were four or five pictures as the test items (see Figure 3.6). Each set of pictures depicted the context that required the use of the past counterfactual hypothetical conditional. The participants were instructed to describe the set of two pictures in one sentence using the verbs presented below each picture. The following four instructions were provided for each set of descriptive pictures:

- (1) Describe the two pictures in one sentence.
- (2) Start each sentence with "If ..."
- (3) Use the verb given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.

In total, there were fourteen items for the target form, which elicited the use of the past hypothetical conditional. The participants were asked to type each of their descriptions in the box presented right below the set of pictures. Since the WPDT was an untimed test, there was no time limit, but they were asked not to spend too much time on one picture description due to the limited laboratory schedule. The scoring procedure of the target linguistic form followed the interlanguage (IL) scoring system used in Shintani et al. (2014) and Shintani (2019), which was

the revised version of Izumi and Bigelow's (2000) interlanguage scoring system. The detailed scoring procedures are provided in the following Coding, Scoring, and Analysis section (see Table 3.4).

Picture 1



- (1) Describe the two pictures in one sentence.
- (2) Start each sentence with "If ..."
- (3) Use the verb given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.



Please write your one-sentence description here.

Figure 3.6. A set of picture cues for the target form in the WPDT

Oral Elicited Imitation Test

Similarly to the WPDT, the oral elicited imitation test (OEIT) was also computerized and was conducted using a PC. The OEIT consists of 21 items: fourteen items for the target linguistic form (i.e., the past hypothetical/counterfactual conditional) and the rest of the seven items as the distractors, which were comprised of various grammatical forms other than the target linguistic form (e.g., relative clauses, infinitive, progressive, causative, and gerund) (see Appendix I for all the test items and the details for each test items).

In previous ISLA studies that used the OEIT, ungrammatical items were used to examine whether the test-takers could unconsciously correct the ungrammaticality and then reconstruct the stimulus sentence using the accurate target linguistic structure. However, ungrammatical items were not used in this study because the target linguistic form was the past hypothetical/counterfactual conditional, which consisted of complex syntactic and semantic properties, and thereby it was predicted to be almost impossible to accurately interpret the meaning of the sentence if the sentence was ungrammatical. A similar issue was reported by Izumi and Bigelow (2000) regarding the use of ungrammatical items in a grammaticality judgment test (GJT). Thus, Izumi and Bigelow (2000) decided to avoid using a GJT in their study because they found that the interpretation of the GJT results of their previous study (Izumi et al., 1999) had become highly problematic. In other words, ungrammatical stimulus sentences may not be interpretable for the test takers and thus may fail to elicit the accurate use of the past hypothetical/counterfactual conditional without providing a specific context.

The main objective of using the OEIT in this study was not to measure the participants' implicit knowledge as was done in many previous ISLA studies (Erlam, 2006; Ellis, 2005; Ellis et al., 2009). Rather, the primary aim of using the OEIT in this study was to measure L2 learners' receptive and productive knowledge in the oral mode using a different type of task from the one of the treatment task. During the treatment tasks, both output and input groups used a set of descriptive pictures during the task phase. Particularly for the output group, they engaged in the text-reconstruction task using picture cues, which was a similar task that they did for the WPDT. From the perspective of the transfer appropriate processing (TAP), which argues that learners can best retrieve the knowledge that they learned in the same/similar context in which the learning took place (Lightbown, 2008), it was valuable to use a measure that did not require the

use of similar sets of skills used in the treatment sessions and then evaluated whether the effectiveness of instruction could be generalizable to other types of task that also required the learners' control over the target linguistic form (Nassaji, 2020).

The format of the OEIT followed the one of Erlam (2006) and Erlam and Akakura (2016), which was conducted through the following procedures: (1) listening to each statement, (2) determining whether the statement is true, false, or not sure for the participants (see Figure 3.7), and then (3) repeating the statement as accurately as possible. Since this was a timed test, the participants had four seconds for the second procedure (i.e., answering their beliefs) and eight seconds for the final repetition (or reconstruction) (see Appendix J for the example display of the OEIT test format).

<i>If I had had money, I would have bought a new computer. (Listening)</i>		
True	False	Not sure

Figure 3.7. An example stimulus-sentence and a belief questionnaire

As for the length of the stimulus statements in the OEIT, medium and long sentence lengths (13-18 syllables) were used in the study based on Yan et al.'s (2016) categorization: short (< 8 syllables), medium (8-15 syllables), and long (> 15 syllables). The vocabulary level of the statements was adjusted so that the participants could achieve a fair comprehension of the test statements (Hu & Nation, 2000). Thus, all the words used in the test were covered within the bands of the most frequent 1K-3K words after excluding proper nouns. In each testing session, the same test was used across the three testing sessions from the pretest to the delayed posttest, but each test was administered by changing the orders of the test items. The same scoring procedures as the one of the WPDT were employed for the scoring of the OEIT performances (see Table 3.4).

Coding, Scoring, and Analysis

Eye-Tracking

Out of a full spectrum of eye-movement measures, both early and late measures (i.e., FPRT and RRT) were analyzed as indications of how much the participants paid attention to each target linguistic exemplar while reading the texts. As it is known that individuals differ markedly in their length and patterns of eye movements (Rayner, 1998, 2009), gain fixation durations on both FPRT and RRT were calculated by subtracting the fixation duration (FPRT and RRT during the first reading [Input 1]) from those of the second reading (Input 2), respectively, both of which were also used for the analysis.

First, both descriptive statistics and their visual counterparts (e.g., boxplots) were inspected. The eye-tracking results on Text 1 and Text 4 and each of these FPRT and RRT results were analyzed separately using *R* version 4.1.1 (R Core Team, 2021). To interpret that results, 95% confidence intervals (CIs) and effect sizes of the between-group and within-group comparisons were also used to analyze the eye-tracking results. Particularly, effect sizes were interpreted based on Plonsky and Oswald's (2014) guidelines for interpreting effect sizes in applied linguistics studies. According to their guidelines, the following interpretation guidelines were suggested: small-ish ($d = .40$), medium-ish ($d = .70$) and large-ish ($d = 1.00$) for the mean difference between groups and small-ish ($d = .60$), medium-ish ($d = 1.00$) and large-ish ($d = 1.40$) for the mean difference within groups (see Plonsky & Oswald, 2014).

Then, the differences in the FPRT and the RRT were compared between the instructional treatment groups (i.e., the Input Only, the Oral Output, and the Written Output groups). After checking the required assumptions, the data were submitted to mixed-design analyses of variance (ANOVAs) using one between-subjects (Group) and one within-subjects (Time) factors. The

between-subject factor (Group) had three levels: the Input Only, the Oral Output, and the Written Output groups, and the within-subject factor (Time) had two levels: the first reading (Input 1) and the second reading (Input 2). Both the FPRT and the RRT gains from the first reading to the second reading were used to examine whether or not they would correlate and predict the gains on the grammatical knowledge measures (the OEIT and the WPDT).

Scoring Procedures of the Past Counterfactual/Hypothetical Conditional

The target linguistic form (the English past hypothetical/counterfactual conditional) is a complex form consisting of the following seven features in the *if* clause and the main clause: (1) the perfect aspect, (2) the past tense, (3) the past participle (PP) form, (4) the modal in the past tense, (5) the perfect aspect, (6) the modal form, and (7) the PP form. Therefore, the target-like (TL) scoring system may miss some signs of learning facilitated through the treatment tasks. To address this point, the participants' test-performances were scored following the interlanguage (IL) scoring system used by Shintani et al. (2014) and Shintani (2019), which was a revised version of the scoring procedure employed by Izumi et al. (1999). As shown in Table 3.4, *if* clause and the main clause were scored separately and the points were accumulated based on the accuracy of the participants' encodings. The possible maximum score for one test item was five points (two points for the *if* clause and three points for the main clause).

Table 3.4. *Scoring system for the past counterfactual hypothetical conditional*

<i>Clause</i>	<i>Criteria</i>	<i>Features</i>	<i>Components</i>	<i>Point</i>
<i>if</i> clause	1	the perfect aspect	<i>have</i> (aux) + verb	1
	2	the past tense	<i>had</i>	0.5
	3	the past participle form	correct PP	0.5
Main clause	4	the modal in the past tense	past modal	1
	5	the perfect aspect	<i>have</i> (aux) + verb	1
	6	the modal form	correct form of <i>have</i> (aux)	0.5
	7	the past participle form	correct form of PP	0.5
Total possible				5

Pretest-Posttest Performances (OEIT and WPDT)

After inspecting data both descriptively and graphically, the participants' test-performances on the OEIT and the WPDT were separately analyzed through descriptive and inferential statistics using *R* version 4.1.1 (R Core Team, 2021). To interpret the results, 95% confidence intervals (CIs) and effect sizes of the group comparisons were also used to analyze these grammar developmental test results.

After checking the required assumptions (see Appendix K), the learners' test-performances on the OEIT and the WPDT were separately analyzed through a series of mixed-design ANOVAs with two independent variables: one between-subjects (Group) factor with three levels: the Input Only, the Oral Output, and the Written Output, and one within-subjects (Time) factor with three levels: the pretest, the immediate posttest, and the delayed posttests. To examine the relationships between the eye-tracking results and these grammar developmental tests (the OEIT and the WPDT), the gain scores from the pretest to the delayed posttest for each grammar test (the OEIT and the WPDT) were used as an outcome variable of the hierarchical multiple regressions.

Piloting Study

A pilot study was conducted on a small scale for the purpose of refining and revising the research instruments and procedures and finding out potential methodological considerations and weaknesses before finalizing the current research design. Specifically, the following considerations were the primary aims of doing the piloting study: (1) checking out the total time needed for the treatment sessions, testing, and eye-tracking preparations (e.g., providing instruction, working on calibration and validations); (2) revising and finalizing the developmental test items (both the OEIT and the WPDT); (3) determining the right participants'

proficiency levels; (4) checking the quality of the developmental test items through item analyses; and (5) familiarizing myself to the entire procedures of the study.

Pilot Participants

The participants for the piloting treatment sessions were nine international graduate students with varying English proficiency levels (Input Only [$n = 3$], Oral Output [$n = 3$], and Written Output [$n = 3$]). Although these nine participants worked on half of the entire instructional treatment materials and both grammar developmental tests, additional eight pilot participants with various backgrounds with varying English proficiency levels also worked on the developmental tests without engaging in the pilot treatment sessions. The background information of these pilot participants is summarized in Table 3.5.

Table 3.5. *Background information of all the pilot participants*

<i>ID</i>	<i>Group</i>	<i>Proficiency</i>	<i>Nationality</i>	<i>L1</i>	<i>LOR</i>	<i>Notes</i>
P1	Input Only	iBT113	Egypt	Arabic	8 ys	SLS Ph.D. Student
P2	W_Output	iBT98	Korea	Korean	3 ys	EFL teacher
P3	Input Only	PBT677	Japan	Japanese	10+ ys	Japanese instructor
P4	O_Output	iBT105	China	Chinese	11 ys	SLS Ph.D. Student
P5	W_Output	iBT88	Japan	Japanese	5 ys	Japanese instructor
P6	O_Output	PBT600	Thailand	Thai	10+ ys	Received PhD in US
P7	O_Output	iBT110	China	Chinese	1.5 ys	Linguistics Ph.D. Student
P8	Input Only	iBT119	China	Cantonese	6 ys	SLS Ph.D. Student
P9	W_Output	iBT100	Vietnam	Vietnamese	2 ys	MA TESOL student
P10	N/A	IELTS7.0	Japan	Japanese	0 yr	EFL university student
P11	N/A	iBT85	Tanzania	Haya	3 ys	MA TESOL student
P12	N/A	iBT112	China	Chinese	5 ys	Linguistics Ph.D. Student
P13	N/A	iBT115	Hong Kong	Chinese	NA	SLS Ph.D. Student
P14	N/A	ITP520	Japan	Japanese	0.5 ys	ESL student
P15	N/A	iBT86	Japan	Japanese	4 ys	ESL student
P16	N/A	iBT96	Taiwan	Chinese	3.5 ys	EFL teacher
P17	N/A	IELTS6.0	China	Chinese	1.5 yr	Received MA in UK

Notes.

SLS = Second Language Studies Ph.D. Program at Michigan State University

Materials and Procedures

For the piloting sessions, the same materials and procedures described above were administered. However, half of the materials were used for the pilot treatment sessions focusing only on the eye-tracking parts (i.e., Texts 1 and 4). Similarly, the testing session was also shortened and was conducted only once for each pilot participant. Hence, there were no posttest sessions in the piloting phase. All the changes after the iteration of the pilot study are discussed below after the presentation of the pilot results.

Pilot Results

In this section, the results of the pilot study for both the eye-tracking measures and the grammar developmental tests are presented. As for the eye-tracking part, the results of the three different measures were examined: the total reading time (TRT), the FPRT, and the RRT. The TRT is the combination of both the early and the late measures (FPRT and RRT). However, the TRT was dropped from the eye-tracking measure of the current study after the piloting phase due to the differential tendencies observed between the results of the FPRT and the RRT in the piloting results.

After the eye-tracking results, the results of both grammar developmental tests (i.e., the OEIT and the WPDT) at this piloting stage are presented with their test reliability.

Eye-Tracking (Piloting)

The results of the eye-tracking (TRT, FPRT, and RRT) on Text 1 are presented in Figures 3.8, 3.9, and 3.10, respectively. Each pilot participant's eye-fixation duration for each AOI is represented as each dot. Based on the results of the TRT (see Figure 3.8), it seemed that P2 and P9 substantially increased their TRT in their second reading after engaging in the written output and P5 indicated a slight increase in TRT in in the second reading. In contrast, other participants

of both the Input Only and the Oral Output group did not show such increases in their second reading. Particularly, the result of the Input Only group showed decreases from the first reading to the second reading.

One unexpected result was that some participants demonstrated longer eye-fixation durations from the first reading. Even the participants in the Input Only group indicated higher TRT, FPRT, and RRT from the phase of the first reading. Particularly, P1 commented that he clearly remembered that he tried to compare what he was not able to perform accurately during the testing session, which was conducted right before the pilot treatment session, and the right grammatical form while he was reading the text for the first time. Therefore, it seemed those test performances could potentially affect their reading behaviors and their processing of the target form if the pretesting were conducted right before the treatment session.

For closer examinations of the results of the early and late measures, the results of both the FPRT and the RRT were separately examined. As for the FPRT (see Figure 3.9), the results indicated that the participants generally spent less time in the second reading regardless of their different instructional conditions. Interestingly, P9 was the only participant that showed an increase in the FPRT in the second reading out of the Written Output group. P8 also showed a slight increase based on the mean FPRT in the second reading but the distribution of the FPRT for each AOI did not show much increase based on the visual inspection. However, different tendencies were observed in the results of the RRT. As shown in Figure 3.10, all the pilot participants in the Written Output group demonstrated increased RRT whereas it seemed that two participants in the Input Only group (P1 and P3) did not re-read the AOIs in the second reading. The Oral Output group also showed a decreased RRT in the second reading. Therefore, clear contrasting tendencies were identified only on the results of the RRT between the Written

Output group and the Input Only group whereas the results of the FPRT generally showed a relatively similar tendency across the groups showing slightly decreased FPRT in the second reading.

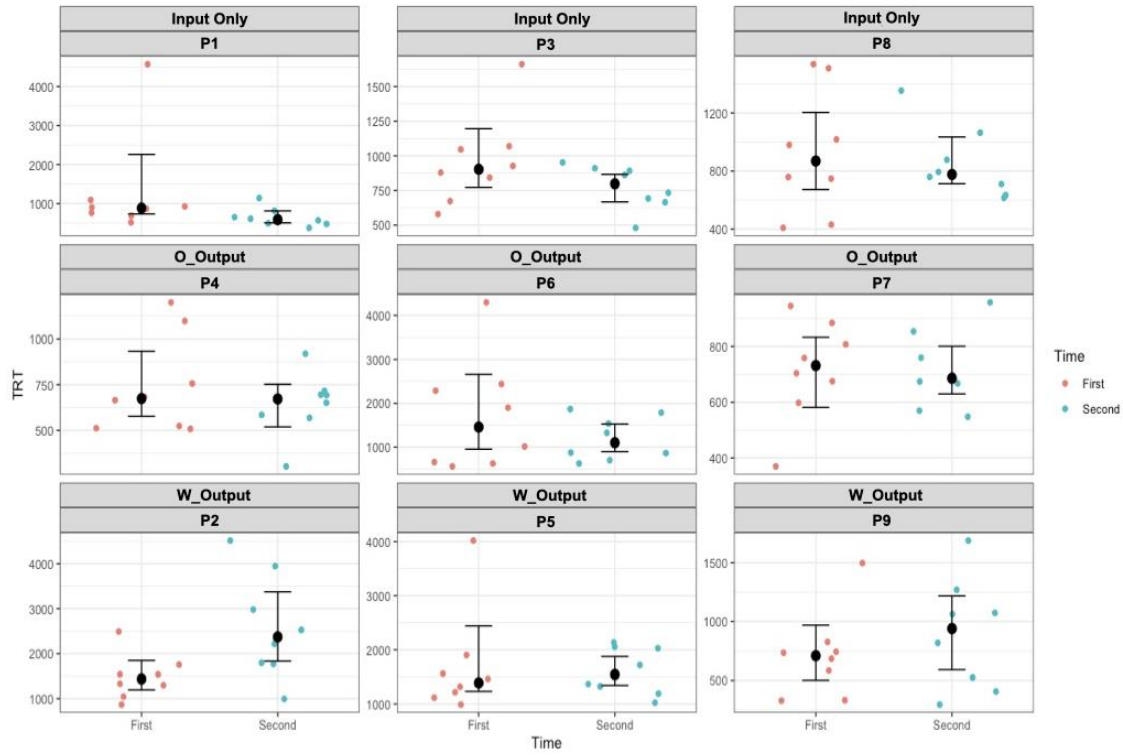


Figure 3.8. Pilot results of the TRT for Text 1

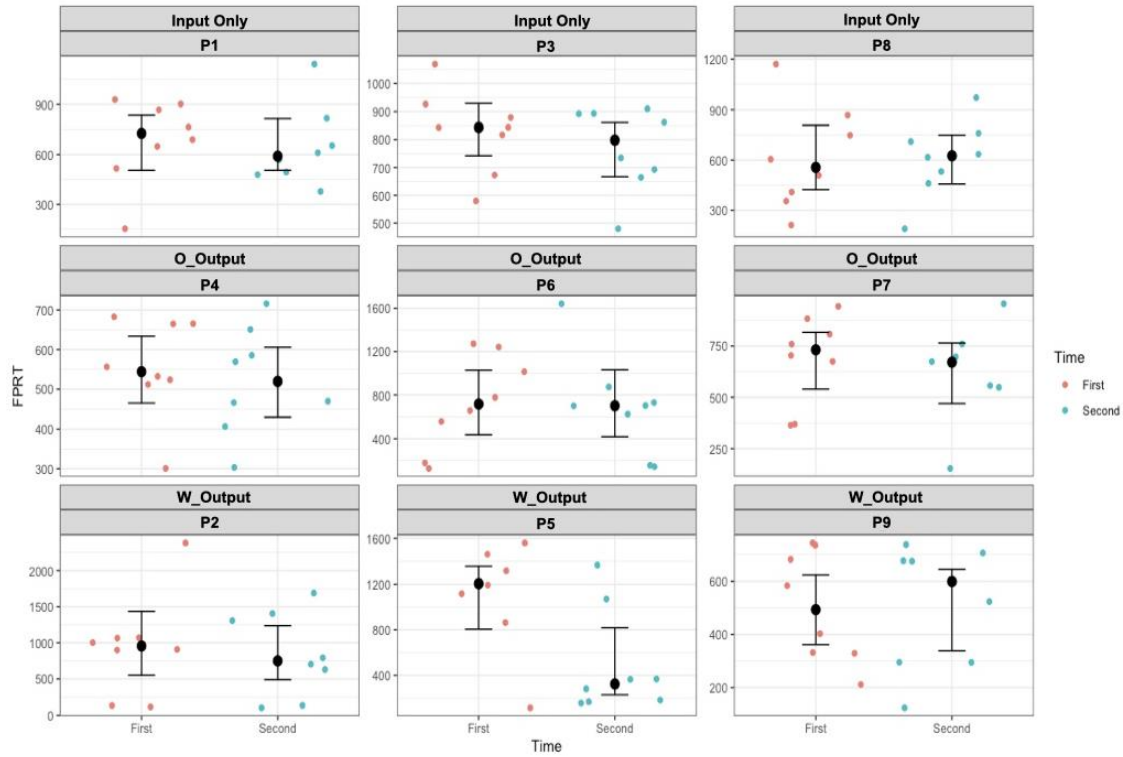


Figure 3.9. Pilot results of the FPRT for Text 1

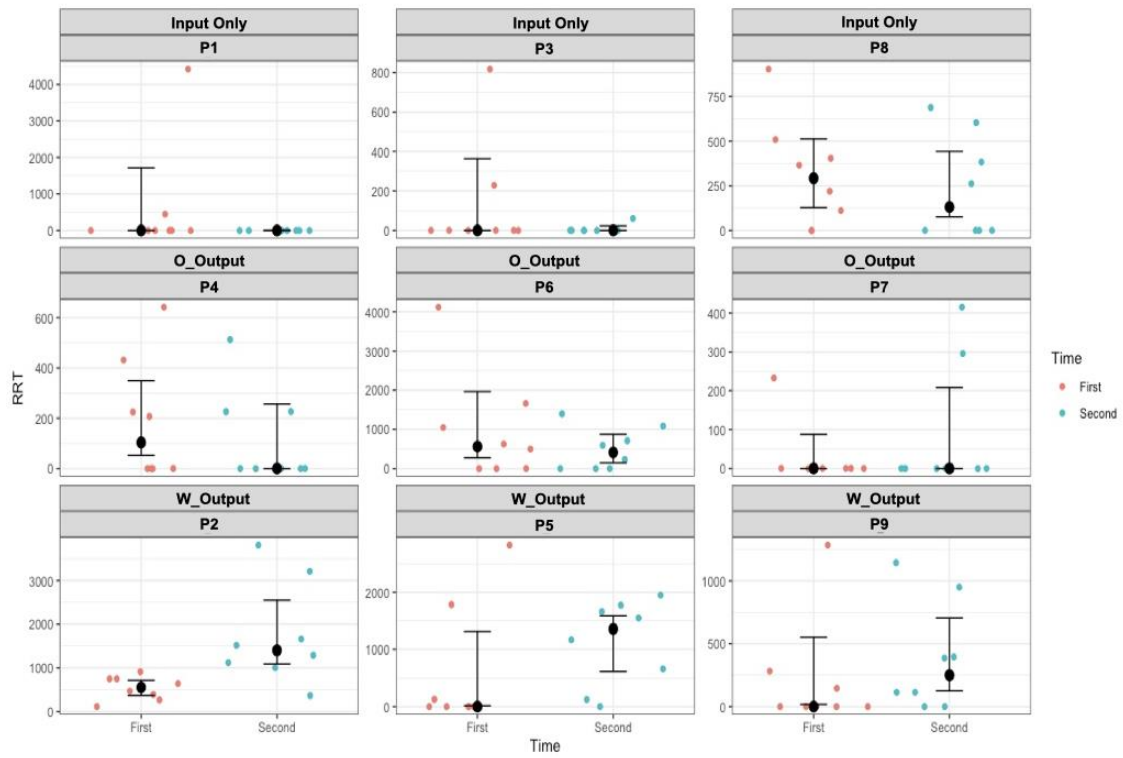


Figure 3.10. Pilot results of the RRT for Text 1

Since the second and the third texts (Text 2 and Text 3) were skipped during the piloting phase, this section presents the eye-tracking results of Text 4, which was arranged to be introduced as the final reading passage during the treatment session. As shown in Figure 3.11, somewhat different tendencies were observed in the results of the TRT for Text 4. The clear opposite patterns between the Written Output group and the Oral Output group were not shown even though P2 and P9 demonstrated slight increases in the TRT during the second reading and the Input Only group showed similar slight decreases from the first reading to the second reading. As for the FPRT for Text 4 (see Figure 3.12), P2 showed a substantial increase in the FPRT during the second reading all the participants generally spent less time or equal time in the second reading, showing slight decreases or equal length of fixation durations for their FPRT. Figure 3.13 shows the results of the RRT for Text 4. Interestingly, both output groups (Written Output and Oral Output groups) did not demonstrate increased RRT for Text 4. The Input Only group showed similar tendencies as the ones of Text 1, indicating very little re-reading or substantial decrease from the first reading to the second reading.

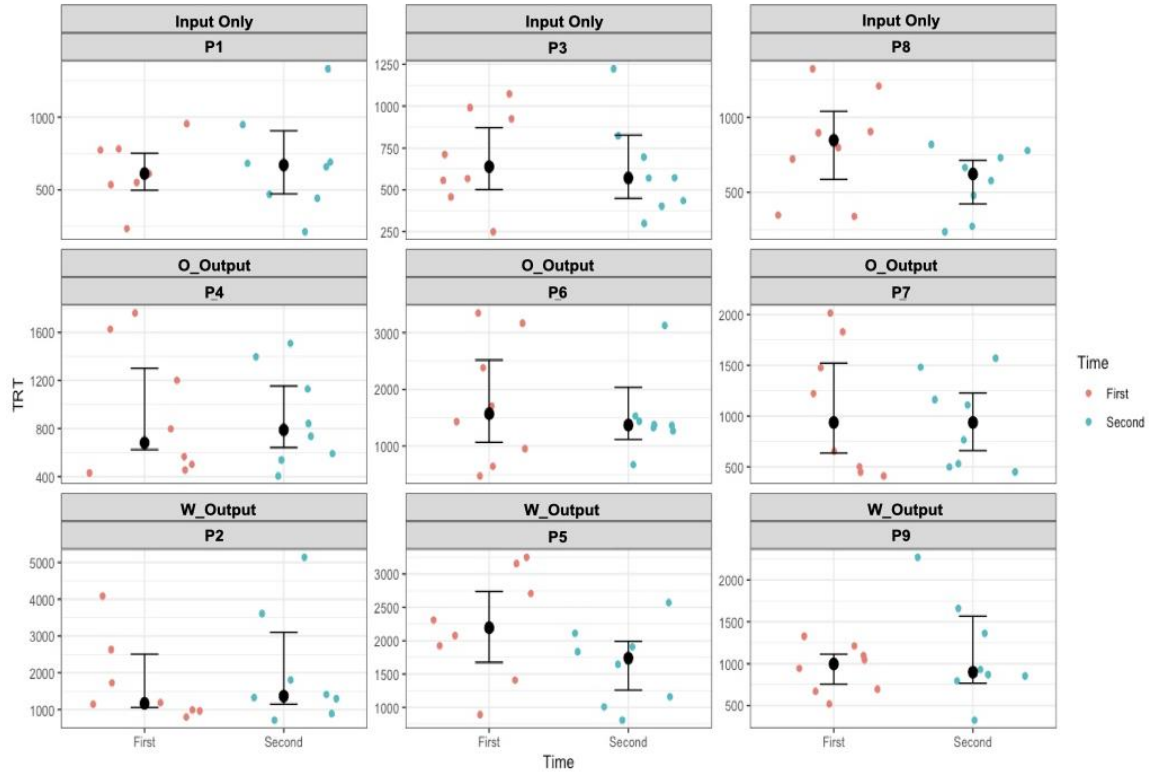


Figure 3.11. Pilot results of the TRT for Text 4

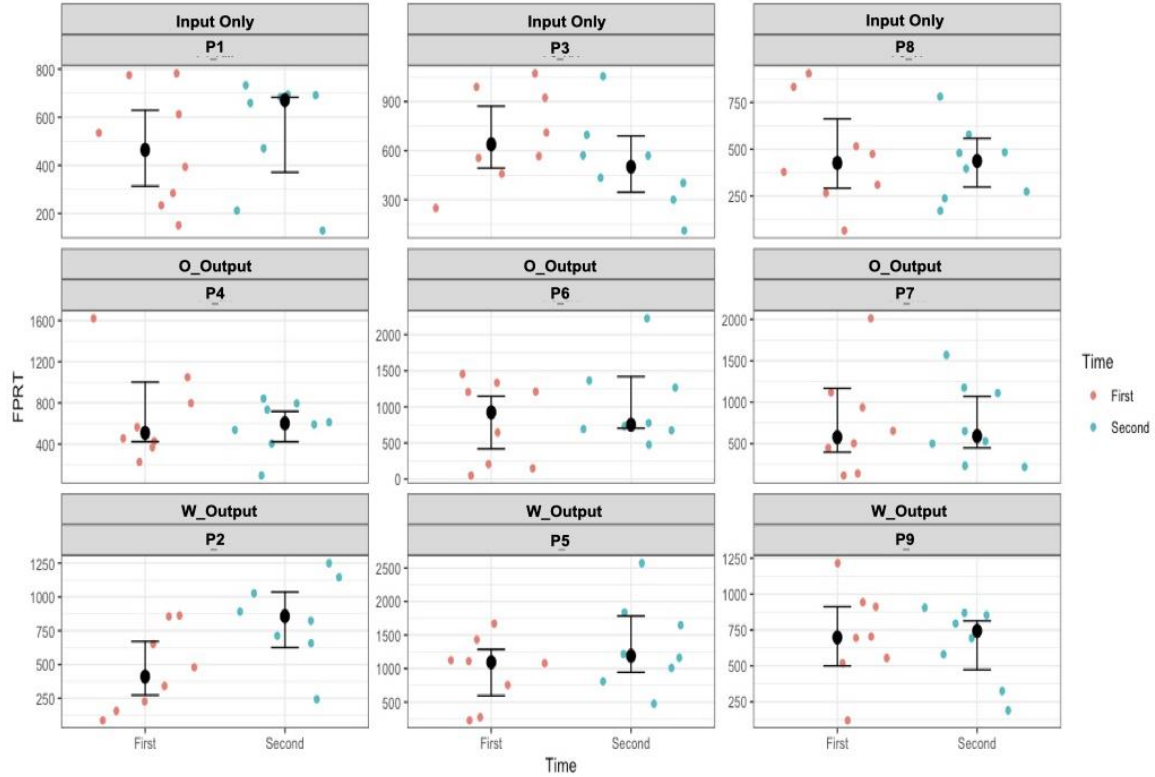


Figure 3.12. Pilot results of the FPRT for Text 4

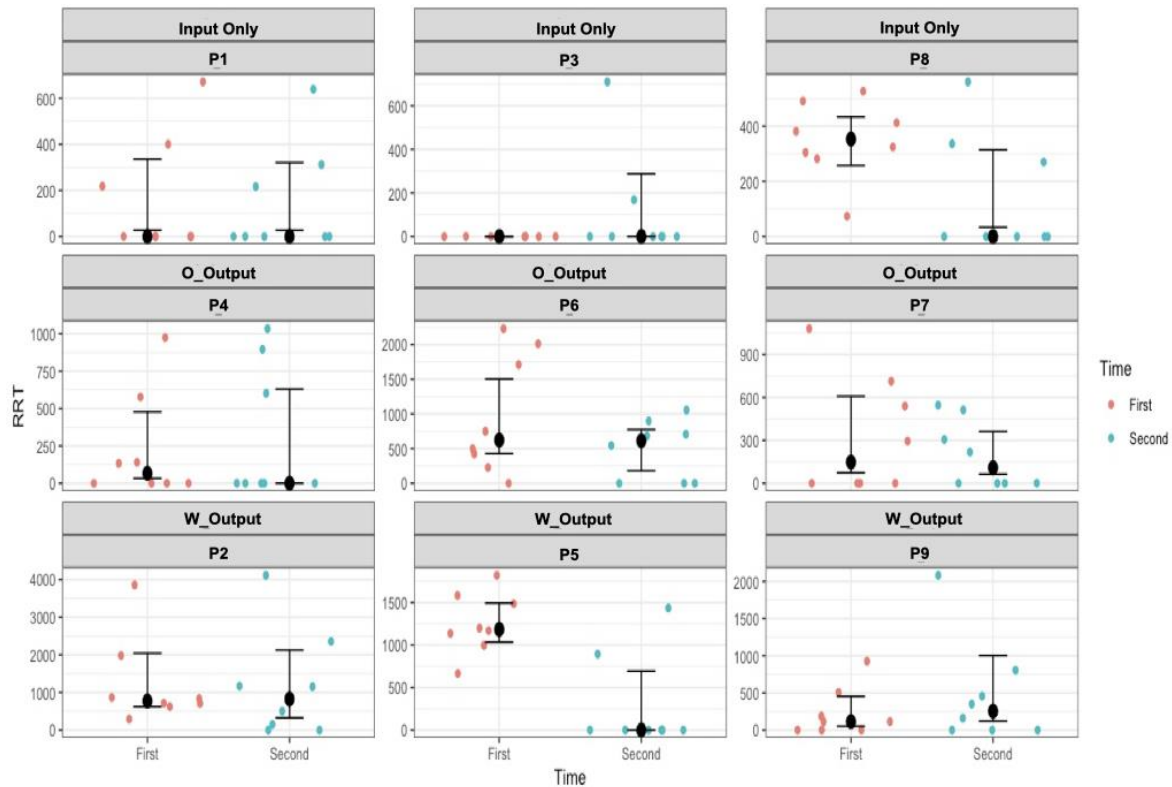


Figure 3.13. Pilot results of the RRT for Text 4

L2 Knowledge Measures (Piloting)

As described above, the main purposes of piloting these grammar developmental tests were the following: checking, revising, and finalizing the developmental test items, determining the right participants' proficiency levels for the target grammatical form, task materials, and test items, and checking the reliability of the test items for each developmental test (the OEIT and the WPDT) before conducting the current study.

As shown in Table 3.6, the pilot participants whose proficiency is higher than B2 (e.g., TOEFL iBT 95 or more; IELTS 7.0 or more) demonstrated ceiling effects showing more than 90% accuracy scores on both tests. Also, EFL teachers and MATESOL students achieved relatively higher scores, especially on the WPDT, which did not require the test-takers' spontaneous automatized knowledge during the test. Based on these results, it can be suggested that learners whose proficiency level can be categorized as C1 or more in CEFL and those who

have English teaching backgrounds should not be included as the participants of the study because their data would be likely to contribute to ceiling effects on the developmental tests. On the other hand, the participants whose proficiency level was B2 in CEFL (e.g., TOEFL iBT 72-95; IELTS 5.5-6.5) appeared to be the right proficiency level for the target linguistic form and the format of the testing materials.

Table 3.6. *Results of both grammar developmental tests (OEIT and WPDT) and the pilot participants' background information*

<i>ID</i>	<i>Proficiency</i>	<i>LOR</i>	<i>Background Information</i>	<i>OEIT</i>	<i>WPDT</i>
P1	iBT113	8 ys	SLS Ph.D. Student	94.00%	96.43%
P2	iBT98	3 ys	EFL teacher	34.00%	99.29%
P3	PBT677	10+ ys	Japanese instructor	55.00%	89.29%
P4	iBT105	11 ys	SLS Ph.D. Student	90.00%	94.29%
P5	iBT88	5 ys	Japanese instructor	33.50%	46.43%
P6	PBT600	10+ ys	Received PhD in US	29.50%	62.86%
P7	iBT110	1.5 ys	Linguistics Ph.D. Student	80.00%	99.29%
P8	iBT119	6 ys	SLS Ph.D. Student	96.00%	99.29%
P9	iBT100	2 ys	MA TESOL student	78.50%	100.00%
P10	IELTS7.0	0 yr	EFL university student	41.00%	95.71%
P11	iBT85	3 ys	MA TESOL student	89.50%	69.29%
P12	iBT112	5 ys	Linguistics Ph.D. Student	82.50%	96.43%
P13	iBT115	NA	SLS Ph.D. Student	88.00%	80.00%
P14	ITP520	0.5 ys	ESL student	11.50%	40.71%
P15	iBT86	4 ys	ESL student	45.50%	40.00%
P16	iBT96	3.5 ys	EFL teacher	78.50%	100.00%
P17	IELTS6.0	1.5 yr	Received MA in UK	2.00%	20.00%

Notes.

SLS = Second Language Studies Ph.D. Program at Michigan State University

To examine the quality of the two grammar developmental tests (i.e., the OEIT and the WPDT), item analysis was carried out by calculating item facility (IF) value, item discrimination (ID) value, and internal consistency. On the results of the items analysis on the OEIT, three items indicated low correlations between the corrected items and the total scores ($r = .65$, $r = .57$, and $r = .65$) out of 20 provisional items. Therefore, these items were removed from the final test items

(14 items in total). After removing these items, a high coefficient of consistency was obtained for the results of the OEIT, $\alpha = 0.97$. As for the WPDT test, all the items were acceptable based on the results of the item analysis. The internal consistency of the WPDT was $\alpha = 0.97$, showing very high reliability. Based on the results of these calculations, the provisional developmental tests and the design were revised.

Revisions Based on the Pilot Study

After conducting a small-scale pilot study, several important methodological issues were identified, and then the design of the study and the materials were revised. Based on these methodological considerations, the following changes and revisions were made to finalize the design, the procedures, and the materials of the current study.

First of all, the timing of the pretesting was changed. As described above, the learners' pretest performances could potentially influence their reading behaviors and processing during the treatment sessions if the pretesting was conducted right before the treatment sessions. As P1 commented, learners could remember the test items and then direct their attention to the form that they struggled with during the treatment session, which could function as another enhancement of the target linguistic form even without working on the output tasks. Therefore, the pretesting and the first treatment session were conducted on different dates and the time gap between the pretest and the first treatment session was set to four to seven days so that the participants would not clearly remember their test performances any longer at the point of the treatment sessions.

Second, an additional eye-tracking practice session was added to the research procedures right before the pretesting. During the piloting phase, it turned out that the calibration and validation processes of the eye-tracking took much more time than scheduled for some

participants. These issues were often related to the participants' unfamiliarity with the calibration and validation processes, reading a passage with eye-tracking, and their eyesight. Also, eye-tracking task did not go smoothly with some specific participants, especially those who were wearing thick glasses or strongly corrected eye-sights with contact lenses. In order to familiarize the participants with the procedures of the eye-tracking tasks and also to cancel the data collection sessions with participants whose eye-movements could not be tracked successfully even before starting the pretesting, an additional eye-tracking practice session was added to the beginning of the data collection sessions. By so doing, the participants could also become familiar with working on the reading task while their eye-movements were tracked during the treatment sessions.

Third, the procedures for the instructional treatments were revised. Especially for the written output group, it took more than 60 minutes to complete one treatment session if learners were required to work on two output trials with the following procedures: the first reading (First Input) → text reconstruction output (First Output) → the second reading (Second Input) → text reconstruction output (Second Output). Since the primary objective was to test the noticing-triggering function of output by comparing the eye-fixation durations on the AOIs between the first reading and the second reading, the second output trial was removed from the research procedures of the current study.

Fourth, based on the results of the eye-tracking measures, TRT was dropped from the eye-tracking (noticing) measures because the early and late measures (i.e., the FPRT and the RRT) showed somewhat different tendencies from each other. Therefore, it was possible that the TRT could obscure the subtle differences indicated between the early and late measures and then

prevent the accurate interpretations of differential tendencies demonstrated by each instructional group.

Finally, the participant recruiting criteria were determined based on the results of the pilot testing and the pilot participants' treatment-task performances. Based on the initial research proposal, the targeted participants were all international undergraduate and graduate students at Michigan State University but both grammar test results indicated that highly proficient international students and those who have English-teaching backgrounds seemed to possess good control over the target linguistic form. Hence, the targeted participants' proficiency level was determined to be B2 level in CEFL without having any English-teaching experiences or TESOL backgrounds.

These were the major methodological modifications made after conducting the pilot study. As shown in the results of the eye-tracking, the Written Output group and the Input Only group indicated somewhat opposite tendencies, indicating increased eye-fixation durations for the Written Output group and decreased fixation durations for the Input Only group. This tendency was pronounced in the results of the late measure (i.e., the RRT) than the ones of the early measure (i.e., FPRT). As for the Oral Output group, these clear opposite tendencies were not observed in their pilot results, but they still showed relatively equal rather than decreased fixation durations from the first reading to their second reading. Considering these results, a sign of the noticing-triggering function of output was indicated for both output groups, particularly for the Written Output group, which aligned with the proposed hypotheses described in the previous chapter (Chapter 2).

CHAPTER 4: RESULTS

This chapter presents the results of both the L2 processing measures and the L2 grammar knowledge measures. For the L2 processing measures, the results of the eye-tracking (i.e., first-pass reading time [FPRT]) on both reading texts (Text 1 and Text 4) are presented. After the presentation of the results of the process-oriented measures, the results of both grammar developmental measures (i.e., the oral elicited imitation test [OEIT] and the written picture description test [WPDT]) are presented. For each type of measure (both processing and grammar developmental measures), first, descriptive statistics and their visual counterparts (i.e., boxplots with mean, median, upper and lower quartiles, and maximum and minimum whiskers) are presented. After the analyses of the descriptive statistics and the visual examinations of these, the results of the inferential statistics are provided. For the first four research questions (Research Questions 1, 2, 3, and 4), the results of mixed-design ANOVA and the follow-up analyses are shown to compare group differences across the two reading sessions for the eye-tracking measures (the first and the second reading) and the three grammar developmental testing sessions between groups (the pretest, the immediate posttest, and the delayed posttest). To address the fifth research question (Research Question 5), associations between the results of the L2 processing measures (i.e., the eye-tracking results) and the results of the L2 grammar knowledge measures (i.e., the OEIT and the WPDT) are shown based on the analyses through correlation and multiple regression.

L2 Processing (Noticing) Measures

Eye-Tracking Results on Text 1

Tables 4.1 and 4.5 summarize the descriptive statistics of both the early measure (i.e., the first-pass reading time [FPRT]) and the late measure (i.e., the re-reading time [RRT]) of the eye-

tracking for Text 1, respectively. The visual counterparts of the descriptive results for both measures are also presented as boxplots in Figures 4.1 and 4.2, respectively.

As shown in Table 4.1 and Figure 4.1, both the Oral Output group and the Input Only group spent less time for their FPRT in the second reading (Reading 2) whereas the Written Output group spent more time in their second reading. These tendencies were indicated by the results of the gain values of their FPRT, indicating that only the Written Output group showed positive value on their gain (i.e., increase from the first reading to the second reading) for FPRT, but the Oral Output and Input Only groups declined on FPRT. Comparing their gains on the FPRT between groups based on their 95% confidence intervals (CIs), the 95% CIs of the Written Output group and the Input Only group did not overlap with each other, indicating that there was a significant difference between these two groups.

Table 4.1. *Descriptive Statistics for First-Pass Reading Time (FPRT)*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Reading 1	633.93	635.31	311.38	1082.75	205.48	[554.94, 712.92]
	Reading 2	545.08	530.62	247.25	1022.12	213.33	[463.07, 627.09]
	Gain	-88.85	-112.44	-358.00	587.12	209.00	[-169.19, -8.51]
O_Output	Reading 1	770.33	728.25	344.38	1236.50	233.07	[685.50, 855.16]
	Reading 2	716.35	701.75	215.00	1340.75	246.47	[626.64, 806.06]
	Gain	-53.98	-80.00	-514.88	433.12	205.31	[-128.71, 20.75]
W_Output	Reading 1	679.40	647.56	382.50	1110.88	182.56	[611.78, 747.02]
	Reading 2	753.29	708.12	382.25	1464.62	258.48	[657.54, 849.04]
	Gain	73.89	59.12	-192.50	447.62	187.12	[4.58, 143.20]

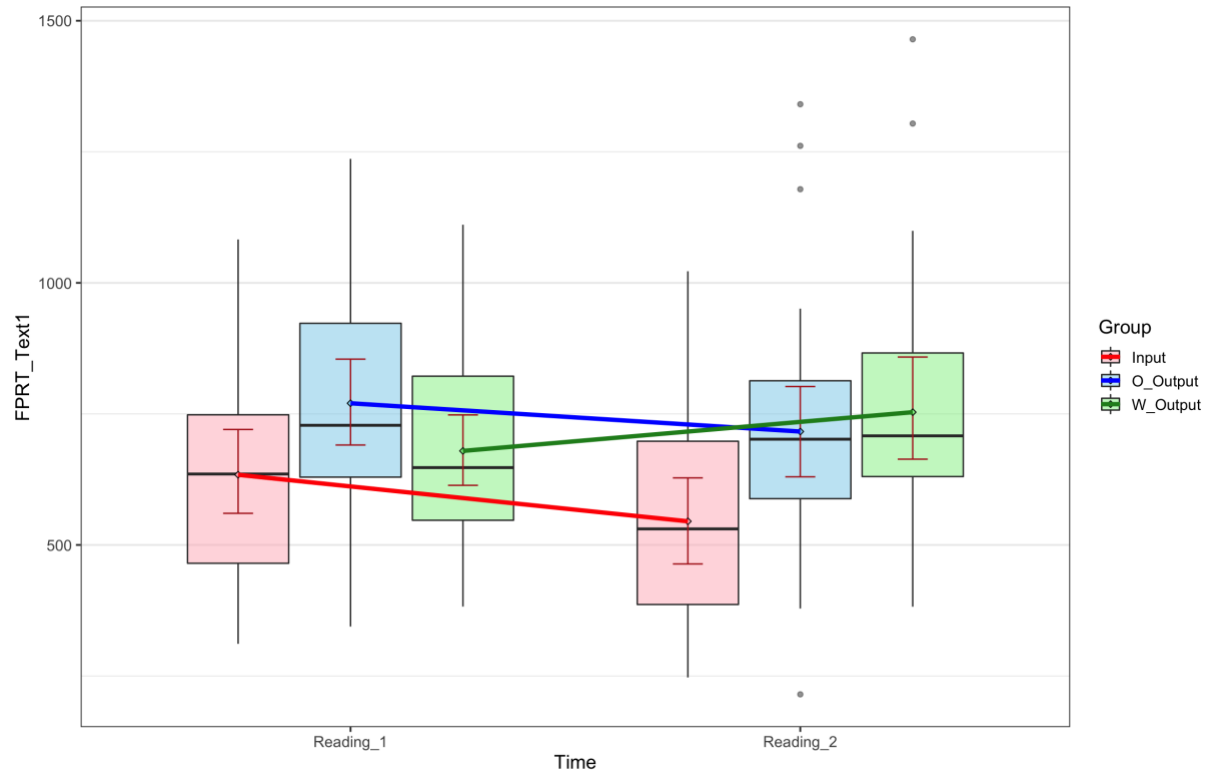


Figure 4.1. Boxplots for first-pass reading time (FPRT) on Text1

These tendencies of the visual inspections were further examined through mixed-design ANOVA and follow-up post-hoc tests. First, one-way ANOVA was performed to compare the group differences on their FPRT at the point of the first reading (Reading 1). The ANOVA results did not indicate significant difference across the groups, $F(2, 80) = 3.08, p = .0512, \eta^2 = 0.07$. The mixed-design ANOVA revealed a significant main effect for Group, $F(2, 80) = 4.46, p < .01, \eta^2 = .08$, and Group x Time interaction, $F(2, 80) = 5.01, p < .01, \eta^2 = .02$, without indicating a significant main effect for Time, $F(2, 80) = 0.98, p = .33, \eta^2 = .002$. The follow-up one-way ANOVA was conducted on the gain FPRT for Text 1 (see Table 4.2). As indicated in Table 4.3, significant group differences between the Input-Only group and the Written Output group, and between the Oral Output group and the Written Output group with medium to large-ish effect sizes.

As for the significant main effect for Time, the opposite tendency was observed between the Input Only group and the Written Output group, indicating a significant increase for the latter and a decrease for the former from the first reading (Reading 1) to the second reading (Reading 2), respectively. A slight decrease was indicated for the Oral Output group's gain FPRT but was not significant (see Table 4.4). Based on these examinations of the results of the descriptive statistics and a series of inferential analyses, the Written Output group showed a somewhat different tendency from the other two groups (the Input Only and the Oral Output group).

Table 4.2. *Follow-up one-way ANOVA on the gain FPRT (Text1)*

<i>Group</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Input_Only	-88.85	5.014	< .01	0.11	Between groups = 2
O_Output	-53.98				Within groups = 80
W_Output	73.89				

Table 4.3. *Pairwise comparisons on the gain FPRT (Text1)*

<i>Group</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only - O_Output	53.98	88.85	.54	0.17
Input_Only - W_Output	54.02	271.46	< .004	0.82
O_Output - W_Output	23.65	232.09	.02	0.65

Table 4.4. *Post-hoc comparisons on each group's FPRT between Reading 1 and Reading 2 (Text1)*

<i>Group</i>	<i>M (Reading 1)</i>	<i>M (Reading 2)</i>	<i>p-value</i>	<i>Cohen's d</i>	<i>95% CIs</i>
Input Only	633.93	545.08	.04	-0.43	[-0.84, 0.02]
O_Output	770.33	716.35	.17	-0.26	[-0.64, 0.11]
W_Output	679.40	753.29	.046	0.39	[0.01, 0.79]

As shown above, only the Written Output group spent more FPRT, which is the indication of the reader's initial stage of sentence processing such as word recognition and lexical access; however, in their second reading, somewhat different tendencies were observed on the 'readers' effortful/non-automatic processing. As shown in Table 4.5 and Figure 4.2, both output groups (Written Output and Oral Output Groups) indicated increased RRT in their second reading with positive gain values for their RRT whereas the one of the Input Only group

decreased in their second reading indicating shorter RRT. As shown in Table 4.5, the 95% CIs on the RRT gains for the Written Output group and the Oral Output group did not overlap with the Input Only group, which indicates that there were significant group differences between both output groups and the Input Only group.

Table 4.5. *Descriptive Statistics for Re-Reading Time (RRT)*

Group	Time	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	95% CIs
Input_Only	Reading 1	374.87	309.38	0.00	1408.00	375.93	[230.36, 519.38]
	Reading 2	166.94	146.81	0.00	698.50	184.09	[96.18, 237.70]
	Gain	-207.93	-91.81	-1263.62	165.25	328.70	[-334.27, -81.59]
O_Output	Reading 1	408.32	207.88	-79.68	1739.00	490.51	[229.78, 586.86]
	Reading 2	709.79	470.62	0.00	2887.62	720.35	[447.62, 971.96]
	Gain	301.47	233.25	-369.25	1840.50	504.85	[117.72, 485.22]
W_Output	Reading 1	258.94	246.81	0.00	667.25	187.28	[189.58, 328.30]
	Reading 2	748.44	413.38	49.00	4280.62	825.18	[442.80, 1054.08]
	Gain	489.51	300.31	-270.88	4105.50	819.45	[185.98, 793.04]

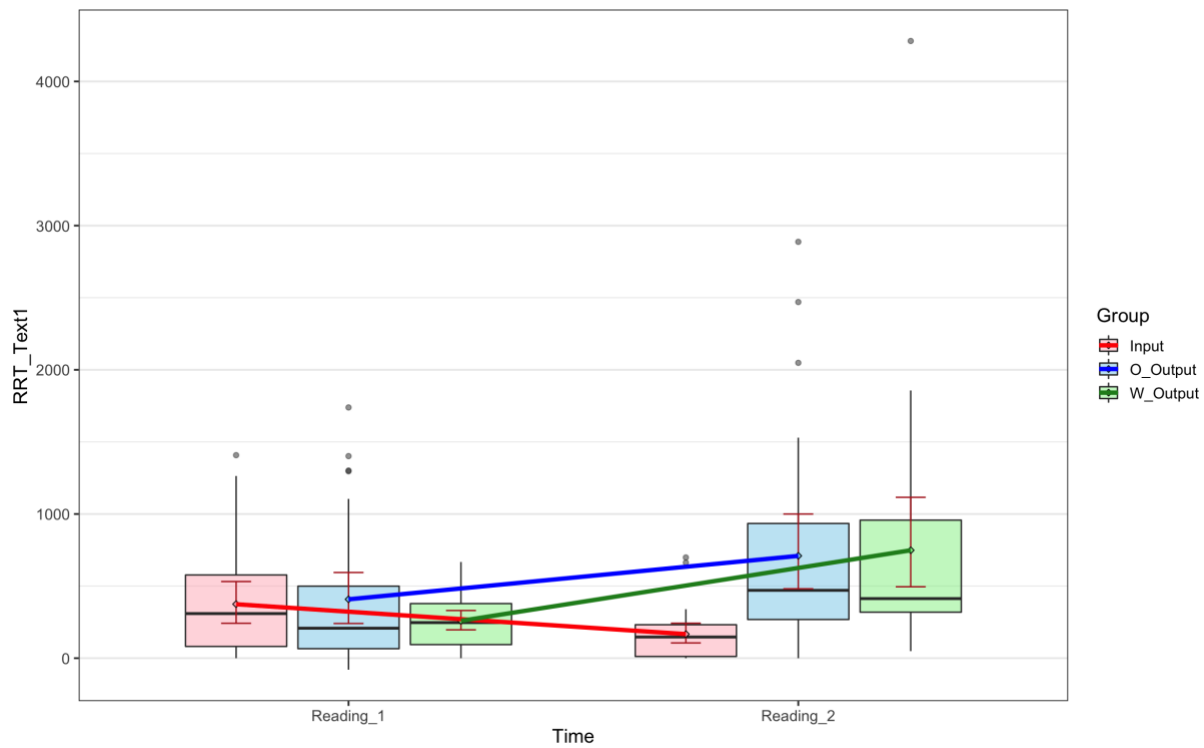


Figure 4.2. Boxplots for re-reading time (RRT) on Text1

To closely examine the group differences and their gains, a mixed-design ANOVA was conducted on the results of the RRT for Text 1. As shown in Table 4.6, no statistically significant group differences were indicated on the results of RRT at the point of the first reading (Reading 1), $F(2, 80) = 1.24, p = .30, \eta^2 = 0.03$. The results of mixed-design ANOVA indicated significant main effects for Group, $F(2, 80) = 3.25, p = .04, \eta^2 = 0.05$, for Time, $F(1, 80) = 10.01, p < .01, \eta^2 = 0.04$, and for Group x Time interaction, $F(2, 80) = 9.97, p < .01, \eta^2 = 0.07$. To specify the differences among the groups, post hoc one-way ANOVA and the pair-wise comparisons with Bonferroni corrections were conducted on each group's gain RRT from the first reading to the second reading. Tables 4.7 and 4.8 show the results of the group differences on the gain RRT, indicating significant differences between the Input Only group and the Oral Output group and also between the Input Only group and the Written Output group with large effect sizes. No significant group difference was indicated between both output groups' gain RRT.

As for the Time effects, a series of post-hoc pairwise comparisons on each group's RRT between the first and second readings were conducted. As shown in Table 4.8, both output groups (Oral Output and Written Output groups) spent significantly more RRT during the second reading than the first reading, on the other hand, the Input Only group spent significantly less RRT during their second reading.

Table 4.6. *Follow-up one-way ANOVA on the gain RRT (Text1)*

<i>Group</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Input_Only	-207.93	9.97	.001	0.20	Between groups = 2
O_Output	301.47				Within groups = 80
W_Output	489.51				

Table 4.7. *Pairwise comparisons on the gain RRT (Text1)*

<i>Group</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input-Only - O_Output	280.71	738.09	< .001	1.18
Input-Only - W_Output	357.24	1037.64	< .001	1.10
O_Output - W_Output	176.65	552.73	.30	0.28

Table 4.8. *Post-hoc comparisons on each group's RRT between Reading 1 and Reading 2 (Text1)*

<i>Group</i>	<i>M (Reading 1)</i>	<i>M (Reading 2)</i>	<i>p-value</i>	<i>Cohen's d</i>	<i>95% CIs</i>
Input Only	374.87	166.94	< .01	-0.63	[-75.17, -340.70]
O_Output	408.32	709.79	< .01	0.60	[109.43, 493.50]
W_Output	258.94	748.44	< .01	0.60	[171.76, 807.26]

Eye-Tracking Results on Text 4

This section presents the eye-tracking results for Text 4, which was the final passage introduced during the treatment sessions (see Table 3.2 and Figure 3.1 for the designs for the treatment sessions and their details in Chapter 3). Tables 4.9 and 13 present the descriptive statistics for the FPRT and the RRT, respectively. Figures 4.3 and 4.4 are the visual counterparts for the early and the late measures. Compared to the eye-tracking results for Text 1, different tendencies were observed for the results of the FPRT and the RRT, respectively, for Text 4.

As shown in Table 4.9 and Figure 4.3, contrasting tendencies were observed between output groups and the Input Only group. Both the Oral Output and the Written Output groups demonstrated slight increases in their FPRT from the first reading to the second reading whereas the Input Only group showed shorter FPRT in their second reading. Based on the comparisons of the 95% CIs for the second reading (Reading 2), the CIs of the Input Only group and both output groups (Oral Output and Written Output) did not overlap with each other, meaning that there were significant group differences on the FPRT of the second reading between the Input Only group and the Oral Output group, and the Input Only group and Written Output group. The gain FPRT overlapped with each other between all three groups since the ranges of the CIs of both output groups indicated negative values for their lower CI values.

Table 4.9. Descriptive Statistics for First-Pass Reading Time (FPRT)

Group	Time	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	95% CIs
Input_Only	Reading 1	632.27	577.56	268.38	1458.88	246.26	[537.60, 726.94]
	Reading 2	559.59	490.93	203.75	1262.12	217.54	[475.98, 643.20]
	Gain	-72.68	-75.32	-423.62	459.00	180.34	[-142.01, -3.35]
O_Output	Reading 1	712.53	657.75	397.88	1011.38	159.76	[654.38, 770.68]
	Reading 2	734.26	700.25	299.75	1411.62	248.58	[643.79, 824.73]
	Gain	21.73	16.88	-436.38	755.62	268.79	[-76.09, 119.55]
W_Output	Reading 1	726.49	769.81	404.50	1182.12	212.06	[647.93, 805.05]
	Reading 2	747.81	690.62	367.62	1257.38	214.93	[668.19, 827.43]
	Gain	21.32	-4.69	-326.50	452.00	192.32	[-49.91, 92.55]

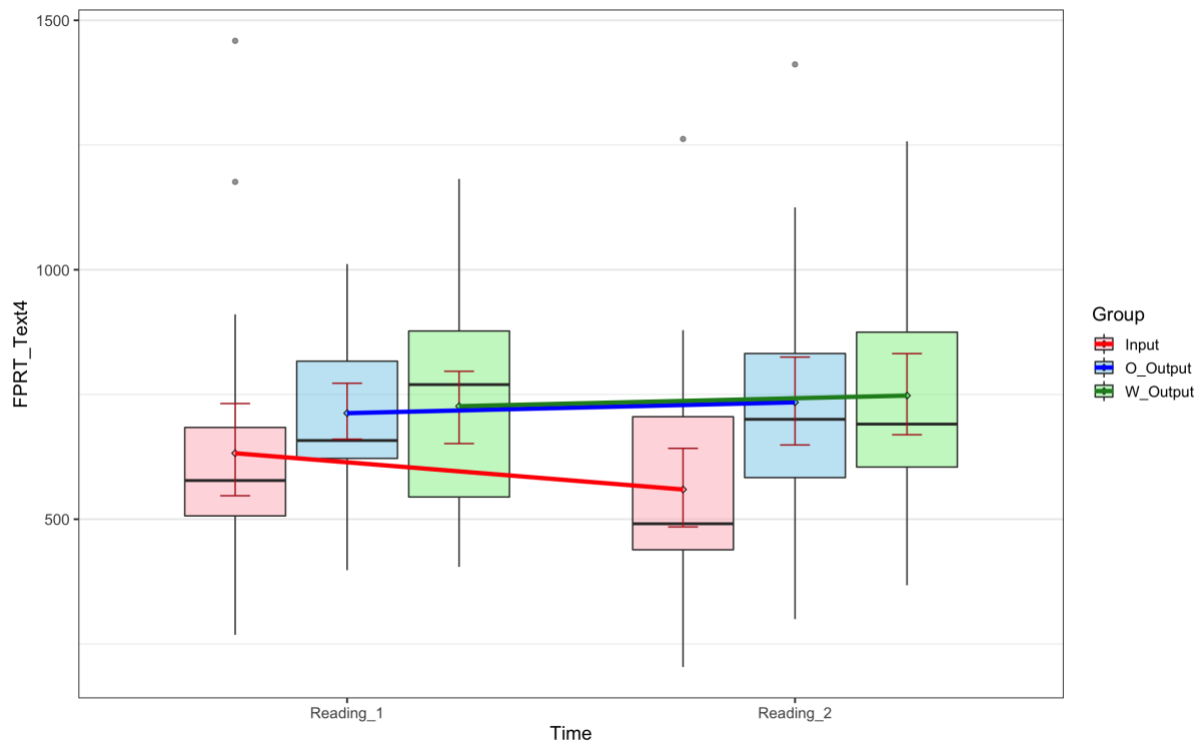


Figure 4.3. Boxplots for first-pass reading time (FPRT) on Text4

These tendencies were confirmed with the results of the mixed-design ANOVA. To examine the group differences and the interaction between Group and Time, one-way ANOVA was conducted to check the group equivalence at the point of the first reading (Reading 1). The ANOVA result did not indicate any significant difference between the groups, $F(2, 80) = 1.60$, $p = .21$, $\eta^2 = 0.04$. The mixed-design ANOVA revealed a significant effect for Group $F(2, 80) =$

4.56, $p = .01$, $\eta^2 = 0.08$, but not for Time, $F(1, 80) = 0.11$, $p = .74$, $\eta^2 = 0.0003$, and Group x Time interaction, $F(2, 80) = 1.65$, $p = .20$, $\eta^2 = 0.01$. Since the main effect was not indicated for Time and Group x Time interaction, the comparisons of the gain FPRT between groups also did not show any significant difference, $F(2, 80) = 1.65$, $p = .20$, $\eta^2 = 0.04$. As for the main effect for Group, significant group differences were revealed for the results of the FPRT of the second reading (Reading 2) (see Tables 4.10 and 4.11). Table 4.12 showed within-group changes from the first reading to the second reading on the results of the FPRT and their effect sizes.

Table 4.10. *Follow-up one-way ANOVA on FPRT of the second reading (Reading 2, Text4)*

<i>Group</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Input_Only	559.59	5.67	< .01	0.12	Between groups = 2
O_Output	734.26				Within groups = 80
W_Output	747.81				

Table 4.11. *Pairwise comparisons on FPRT of the second reading (Reading 2, Text4)*

<i>Group</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only - O_Output	-300.75	-48.60	< .01	0.75
Input_Only - W_Output	-306.44	-69.99	< .01	0.87
O_Output - W_Output	-136.80	109.711	.83	0.06

Table 4.12. *Post-hoc comparisons on each group's FPRT between Reading 1 and Reading 2 (Text4)*

<i>Group</i>	<i>M (Reading 1)</i>	<i>M (Reading 2)</i>	<i>p-value</i>	<i>Cohen's d</i>	<i>95% CIs</i>
Input Only	632.27	559.59	.05	-0.40	[-145.52, 0.16]
O_Output	712.53	734.26	.67	0.08	[-80.51, 123.97]
W_Output	726.49	747.81	.56	0.11	[-53.25, 95.89]

The results of the RRT for Text 4 also indicated somewhat different tendencies from the ones for Text 1. The results of the RRT for Text 1 indicated that both output groups (the Written Output group and the Oral Output group) significantly increased their RRT from the first reading to the second reading while the Input Only group showed a significant decrease in their RRT from the first reading to the second reading. For the reading of Text 4, however, each group's 95% CIs between the first reading and the second reading overlapped with each other, suggesting

that their RRT did not indicate significant differences in time between the first reading and the second reading. As for the 95% CIs for their gain RRT, the ranges of each group's CIs did not overlap with each other. One unique difference from the results of the RRT for Text 1 was that the Oral Output group seemed to spend substantially more RRT from the first reading of Text 4, which was indicated by the non-overlap of the CIs between the Input Only group's RRT on the first reading and the Oral Output group's RRT on their first reading. The tendency of the initial increase in RRT for the first reading was also observed in the Written Output group's RRT, but the ranges of their 95% CIs overlapped with the ones of the Input Only group.

Table 4.13. *Descriptive Statistics for Re-Reading Time (RRT)*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Reading 1	257.52	202.31	0.00	1086.12	261.61	[156.95, 358.09]
	Reading 2	141.19	111.62	0.00	557.62	144.44	[85.66, 196.72]
	Gain	-116.33	-78.31	-602.75	232.89	218.24	[-200.22, -32.44]
O_Output	Reading 1	598.61	573.25	0.00	2337.75	576.81	[388.67, 808.55]
	Reading 2	610.20	449.62	0.00	3900.25	782.82	[325.27, 895.13]
	Gain	11.59	-18.88	-936.80	1562.50	472.64	[-160.44, 183.62]
W_Output	Reading 1	369.40	260.81	0.00	1359.62	354.99	[237.90, 500.90]
	Reading 2	551.60	439.69	25.75	2031.50	483.41	[372.53, 730.67]
	Gain	182.20	44.50	-678.88	1046.50	434.32	[21.32, 343.08]

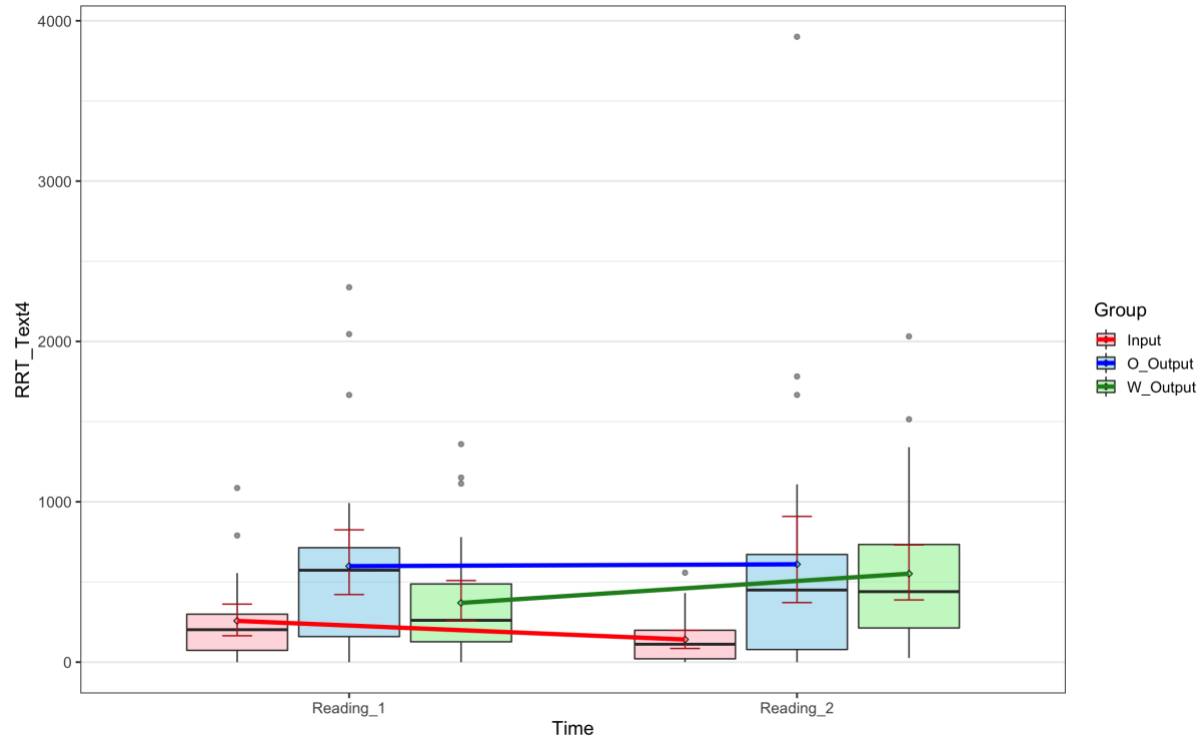


Figure 4.4. Boxplot for re-reading time (RRT) on Text4

For the results of the RRT for Text 4, mixed-design ANOVA was not conducted due to the significant group differences at the point of the first reading (Reading 1) (see Table 4.14). As shown in Table 4.15, the Oral Output group showed significantly higher RRT in the first reading than the ones of the other groups (i.e., the Input Only group and the Written Output group). In other words, the Oral Output group spent significantly more time to re-read the AOIs (or the structures of the target grammatical form) from the first reading than did the other two groups. Due to such group differences at the point of the first reading, analysis of covariance (ANCOVA) was conducted, instead of mixed-design ANOVA, on the results of each group's gain RRT using the RRT of the first reading as a covariate. The results of ANCOVA revealed that the RRT gains indicated significant differences across the groups, $F(2) = 4.09, p = .02$, controlling for the RRT of the first reading, $F(1) = .55, p = .46$ (see Table 14.6). Tukey post hoc tests indicated that there was a significant difference between the gain RRT of the Written

Output group and the one of the Input Only group, whereas a significant difference was not indicated between the gain RRT of Input Only group and the one of the Oral Output group, and between the gain RRT of the Written Output group and the one of the Oral Output group (see Table 4.17). As Table 4.18 shows, the within-group gains on the RRT indicated the opposite tendencies between the Input Only group and the Written Output group.

Table 4.14. *Results of one-way ANOVA on RRT of the first reading (Reading 1, Text4)*

<i>Group</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Input_Only	257.52	4.66	.01	0.10	Between groups = 2
O_Output	598.61				Within groups = 80
W_Output	369.40				

Table 4.15. *Pairwise comparisons on RRT of the first reading (Reading 1, Text4)*

<i>Group</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only - O_Output	-581.13	-101.05	< .01	0.75
Input_Only - W_Output	-281.56	57.79	.19	0.36
O_Output - W_Output	-25.08	483.49	.08	0.36

Table 4.16. *ANCOVA on gain RRT with RRT of the first reading (Reading 1) as a covariate*

	<i>SS</i>	<i>df</i>	<i>F</i>	<i>p-value</i>
RRT4_1	86783	1	0.55	.46
Group	1281940	2	4.09	.02
Residual	12385038	79		

Note. RRT4_1: Re-reading time for the first reading of Text 4, SS: Sum of squares

Table 4.17. *Pairwise comparisons on gain RRT from Reading 1 to Reading 2 (Text4)*

<i>Group</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only - O_Output	-105.97	432.21	.32	0.34
Input_Only - W_Output	51.08	569.06	.01	0.86
O_Output - W_Output	-109.96	403.85	.36	0.38

Table 4.18. *Post-hoc comparisons on each group's RRT between Reading 1 and Reading 2 (Text4)*

<i>Group</i>	<i>M (Reading 1)</i>	<i>M (Reading 2)</i>	<i>p-value</i>	<i>Cohen's d</i>	<i>95% CIs</i>
Input Only	257.52	141.19	.01	-0.53	[-204.47, -28.18]
O_Output	598.61	610.20	.90	0.02	[-0.90, 191.38]
W_Output	369.40	551.60	.04	0.42	[13.79, 350.61]

To summarize the results of the eye-tracking on both the FPRT and the RRT for both texts (Text 1 and 2), the eye-tracking results generally indicated the opposite pattern between the output groups (the Oral Output and the Written Output groups) and the non-output group (the Input Only group). Both output groups showed increased fixation durations from the first reading to the second reading whereas the Input Only group demonstrated consistent decreases of their FPRT and the RRT from the first reading to the second reading. These opposite patterns were more evident in the results of the RRT, especially between the Input Only group and the Written Output groups. The results of the series of mixed-design ANOVAs showed significant Group x Time interactions especially for the first passage (Text 1), confirming these different tendencies between the groups depending on each instructional treatment condition. Also, somewhat different tendencies were observed within each group between Text 1 and Text 4, the possible explanations and the implications are discussed in the following discussion chapter (Chapter 5).

L2 Grammar Knowledge Measures

The previous section presented the eye-tracking results as the process-oriented measures of L2 learning. In this section, the results of both grammar developmental tests (OEIT and WPDT) are presented respectively as the product-oriented measures of L2 learning.

Results of the Oral Elicited Imitation Test (OEIT)

Tables 4.19 and 4.20 present the descriptive statistics of the results of the OEIT in the pretest, the immediate posttest, and the delayed posttest and the gains scores, respectively. The boxplots in Figures 4.5 and 4.6 visually present these results.

As shown in Tables 4.19 and 4.20 and Figures 4.5 and 4.6, the results of the OEIT showed that all the instructional groups similarly improved their posttest performances both in the immediate posttest and the delayed posttest. Although the 95% CIs of the Oral Output group

between the pretest and the immediate posttest did not overlap with each other, the three groups generally showed significant increases from the pretest to the posttests, which were indicated by the non-overlaps of their ranges of the 95% CIs between each testing session. As for the gain scores, the Written Output group showed the highest immediate gains but all the ranges of the 95% CIs between groups overlapped with each other (see Table 4.20 and Figure 4.6), suggesting that all the instructional groups showed similar tendencies of the gains throughout the OEIT sessions.

Table 4.19. *Descriptive Statistics for the OEIT*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Pre	29.90	31.50	2.00	54.50	13.04	[24.88, 34.92]
	Post1	43.02	41.00	10.50	67.50	16.50	[36.67, 49.37]
	Post2	45.79	49.50	14.00	69.50	17.02	[39.24, 52.34]
O_Output	Pre	29.86	30.00	4.00	54.00	15.61	[24.18, 35.54]
	Post1	40.93	42.00	12.00	65.00	16.99	[34.76, 47.10]
	Post2	42.97	47.50	8.00	66.00	16.86	[36.84, 49.10]
W_Output	Pre	29.82	27.00	1.00	64.00	17.62	[23.29, 36.35]
	Post1	45.59	50.00	6.00	67.00	17.47	[39.12, 52.06]
	Post2	44.34	49.25	6.00	69.50	17.05	[38.03, 50.65]

Table 4.20. *Descriptive Statistics for Gain Scores on the OEIT*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Gain1	13.12	11.75	-0.50	28.50	7.90	[10.08, 16.16]
	Gain2	15.88	15.75	-1.00	31.50	8.05	[12.78, 18.98]
O_Output	Gain1	11.07	9.00	-3.50	30.50	7.03	[6.99, 15.15]
	Gain2	13.10	13.50	-10.00	32.50	10.19	[8.69, 17.51]
W_Output	Gain1	15.77	12.75	-2.50	41.50	10.98	[11.69, 19.85]
	Gain2	14.52	14.00	-6.00	42.50	11.92	[10.11, 18.93]

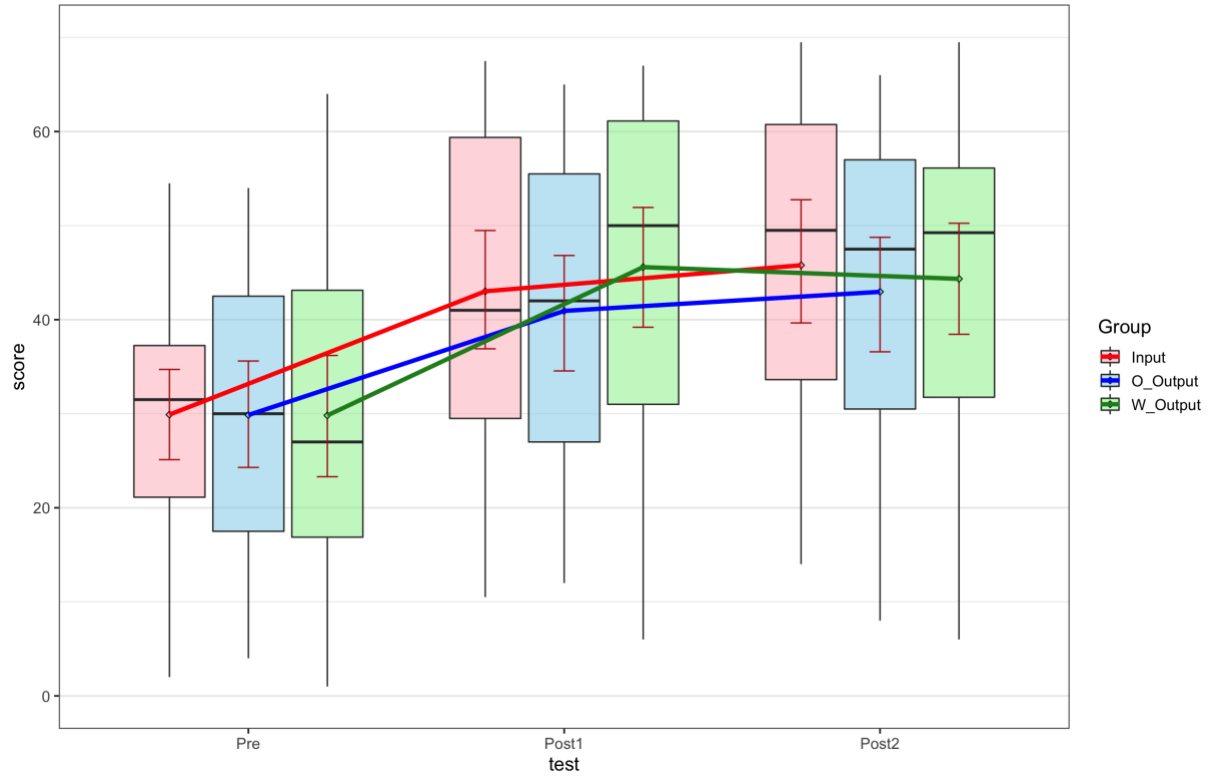


Figure 4.5. Boxplots for the results of the OEIT

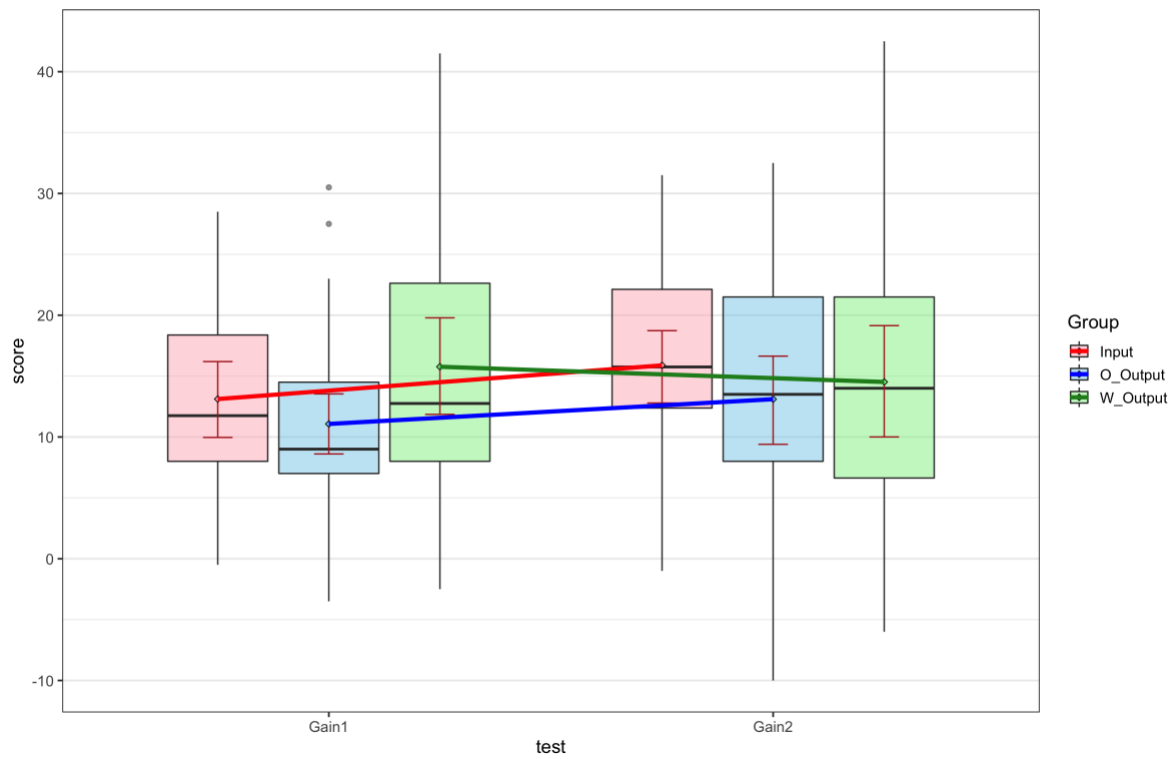


Figure 4.6. Boxplots for the gain scores on the OEIT

To closely examine the general tendencies observed based on the results of the descriptive statistics and the visual inspections, mixed-design ANOVA was conducted with Group as a between-subjects and Time as a within-subject variable. First, one-way ANOVA was performed on each group's pretest scores to examine the group equality at the start of the study. The results of one-way ANOVA indicated no significant group differences at the point of the pretest, $F(2, 80) = 0.0001, p = .099, \eta^2 < 0.001$. The result of the mixed-design ANOVA showed only a significant main effect for Time, $F(2, 160) = 139.24, p < .001, \eta^2 = 0.14$, and no significant effects for Group, $F(2, 80) = 0.13, p = .88, \eta^2 = 0.003$, and the Time x Group interaction, $F(4, 160) = 1.46, p = .22, \eta^2 = 0.003$, meaning that all the instructional groups (i.e., Input Only, Oral Output, and Written Output Groups) improved their OEIT performances similarly as they engaged in the OEIT throughout the testing sessions. As Tables 4.21, 4.22, and 4.23 show, post-hoc repeated measures ANOVA and pair-wise comparisons indicated that all three groups improved significantly in the immediate posttest with large effect sizes and then retained the gains until the delayed posttest. The Input Only group demonstrated steady increase from the pretest to the delayed posttest, also indicating a significant gain from the immediate posttest to the delayed posttest with a small effect size (see Table 4.24).

Table 4.21. Follow-up repeated measures ANOVA on the Input Only group's OEIT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	<i>η^2</i>	<i>df</i>
Pretest	29.90	76.40	< .001	0.17	Between groups = 2
Postt1	43.02				Within groups = 50
Posttest 2	45.79				

Table 4.22. Follow-up repeated measures ANOVA on the Oral Output group's OEIT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Pretest	29.86	40.52	< .001	0.11	Between groups = 2
Postt1	40.93				Within groups = 56
Posttest 2	42.97				

Table 4.23. Follow-up repeated measures ANOVA on the Written Output group's OEIT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Pretest	29.82	39.61	< .001	0.15	Between groups = 2
Postt1	45.59				Within groups = 54
Posttest 2	44.34				

Table 4.24. Pairwise comparisons of time on the OEIT for each group

<i>Group</i>	<i>Test</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only	Pretest - Posttest1	1.08	2.29	< .01	1.66
	Pretest - Posttest2	1.33	2.68	< .01	1.97
	Posttest1- Posttest2	0.20	1.06	.01	0.62
O_Output	Pretest - Posttest1	1.04	2.15	< .01	1.57
	Pretest - Posttest2	0.80	1.81	< .01	1.29
	Posttest1- Posttest2	-0.11	0.64	.51	0.26
W_Output	Pretest - Posttest1	0.91	2.00	< .01	1.44
	Pretest - Posttest2	0.73	1.73	< .01	1.22
	Posttest1- Posttest2	-0.54	0.22	1.00	0.16

Results of the Written Picture Description Test (WPDT)

As the results of the OEIT showed that all the groups significantly improved their posttest performances without showing a Group x Time interaction, a similar tendency was observed in the results of the WPDT. Tables 4.25 and 4.26 present the descriptive statistics of the pretest, the immediate posttest, and the delayed posttest results and the gains scores from the pretest to the posttests, respectively. The visual counterparts are presented in Figures 4.7 and 4.8 as boxplots. Based on these descriptive results, the Written Output group attained the highest posttest scores and gain scores both on the immediate posttest and the delayed posttest among the three groups.

The other two groups also improved their posttest performances. Comparing the ranges of the 95% CIs within group and between groups, the ranges of the 95% CIs of the Written Output group indicated non-overlaps between the ranges of their pretest and immediate posttest and also between their pretest and the delayed posttest while all the ranges of the 95% CIs of the other two groups overlapped with each other within each group (see Table 4.25). Based on these observations, the Written Output group appeared to improve their test performances from the pretest to both posttests more than did the other two groups.

Table 4.25. *Descriptive Statistics for the WPDT*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Pre	33.60	35.75	6.00	70.00	21.05	[25.51, 41.69]
	Post1	44.83	51.75	11.00	70.00	22.26	[36.26, 53.40]
	Post2	47.12	61.25	11.00	70.00	24.06	[37.87, 56.37]
O_Output	Pre	29.43	25.00	1.00	69.50	19.74	[22.26, 36.60]
	Post1	40.45	39.50	12.00	70.00	18.67	[33.65, 47.25]
	Post2	43.17	45.50	1.00	70.00	19.61	[36.04, 50.30]
W_Output	Pre	31.77	31.25	0.00	69.00	19.45	[24.58, 38.96]
	Post1	50.61	58.00	6.00	70.00	21.67	[42.59, 58.63]
	Post2	53.62	67.00	6.00	70.00	22.44	[45.31, 61.93]

Table 4.26. *Descriptive Statistics for Gain Scores on the WPDT*

<i>Group</i>	<i>Time</i>	<i>M</i>	<i>Mdn</i>	<i>Min</i>	<i>Max</i>	<i>SD</i>	<i>95% CIs</i>
Input_Only	Gain1	11.23	9.00	-18.00	53.50	14.27	[5.74, 16.72]
	Gain2	13.52	7.00	-2.00	44.00	14.08	[8.11, 18.93]
O_Output	Gain1	11.02	7.00	-8.50	56.50	13.29	[6.18, 15.86]
	Gain2	13.74	12.00	-5.00	54.00	15.90	[7.96, 19.52]
W_Output	Gain1	18.84	18.50	-1.50	52.50	15.49	[13.10, 24.58]
	Gain2	21.86	19.75	-2.00	52.50	17.34	[15.43, 28.29]

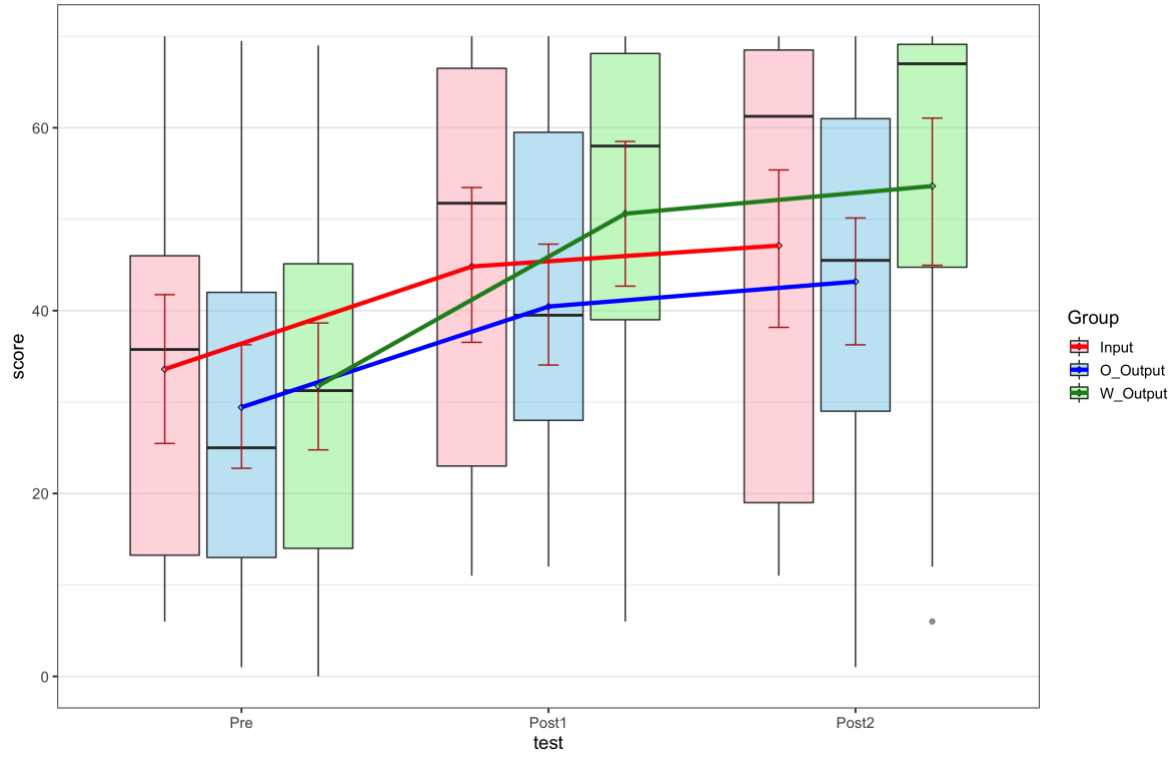


Figure 4.7. Boxplots for the results of the WPDT

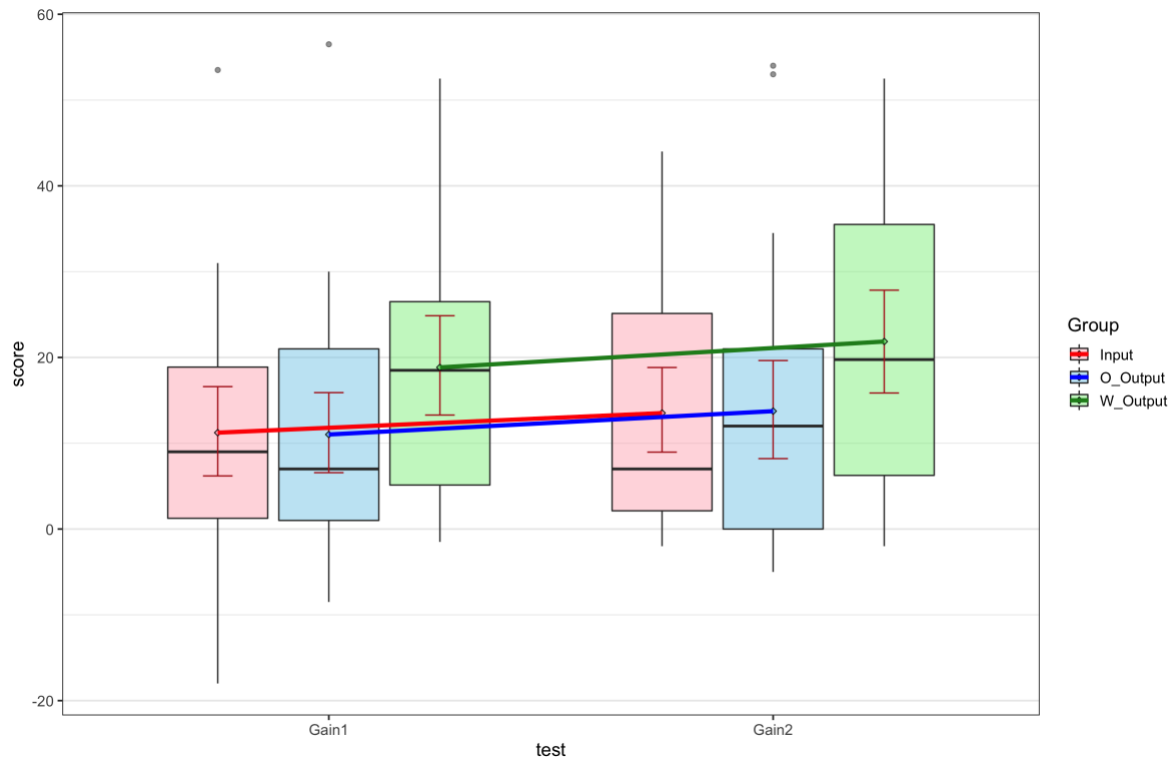


Figure 4.8. Boxplots for the gain scores on the WPDT

To further examine these observations with inferential statistics, mixed-design ANOVA was conducted on the results of the WPDT. First, the group equivalence at the beginning of the study was confirmed through one-way ANOVA on the results of each group's pretest, $F(2, 80) = 0.30, p = .74, \eta^2 = 0.007$. The mixed-design ANOVA showed a significant main effect for Time, $F(2, 160) = 62.78, p < .001, \eta^2 = 0.11$, but not for Group, $F(2, 80) = 1.12, p = .33, \eta^2 = 0.02$, and Group x Time interaction, $F(4, 160) = 1.93, p = .18, \eta^2 = 0.01$. Since a significant main effect was indicated only for Time without any main effects for Group and Group x Time interaction, the ANOVA results suggested that the three groups improved their posttest performances similarly. To closely examine the Time effect for each group, post-hoc repeated measures ANOVA and pair-wise comparisons were conducted (see Tables 4.27, 4.28, and 4.29). The post-hoc results indicated that all three groups demonstrated significant increases in their posttest performances on the WPDT. Based on the effect sizes, the Written Output group showed large effect sizes for their posttest gains (see Table 4.30) and the other two groups showed medium effect sizes for their posttest gains. As shown above, however, these differences in their within-group effect sizes and the differences observed on the results of the descriptive statistics, particularly the highest posttest gains demonstrated by the Written Output group, were not indicated on the ANOVA results.

Table 4.27. Follow-up repeated measures ANOVA on the Input Only group's WPDT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Pretest	33.60	12.76	< .001	0.07	Between groups = 2
Postt1	44.83				Within groups = 50
Posttest 2	47.12				

Table 4.28. Follow-up repeated measures ANOVA on the Oral Output group's WPDT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Pretest	29.43	17.29	< .001	0.09	Between groups = 2
Postt1	40.45				Within groups = 56
Posttest 2	43.17				

Table 4.29. Follow-up repeated measures ANOVA on the Written Output group's WPDT performances from the pretest to the delayed posttest

<i>Time</i>	<i>Mean</i>	<i>F-statistic</i>	<i>p-value</i>	η^2	<i>df</i>
Pretest	31.77	34.89	< .001	0.18	Between groups = 2
Postt1	50.61				Within groups = 54
Posttest 2	53.62				

Table 4.30. *Pairwise comparisons of time on the WPDT for each group*

<i>Group</i>	<i>Test</i>	<i>Lower CIs</i>	<i>Higher CIs</i>	<i>p-value</i>	<i>Cohen's d</i>
Input_Only	Pretest - Posttest1	0.35	1.25	.001	0.79
	Pretest - Posttest2	0.50	1.45	< .001	0.96
	Posttest1- Posttest2	-0.24	0.54	1.00	0.15
O_Output	Pretest - Posttest1	0.41	1.27	< .001	0.83
	Pretest - Posttest2	0.44	1.31	< .001	0.86
	Posttest1- Posttest2	-0.11	0.65	.48	0.27
W_Output	Pretest - Posttest1	0.73	1.73	< .001	1.22
	Pretest - Posttest2	0.77	1.78	< .001	1.26
	Posttest1- Posttest2	-0.12	0.65	.54	0.26

Relationship between L2 Noticing and Grammar Learning

In this section, the relationship between learner noticing that was gauged through the eye-tracking (i.e., both the FPRT and the RRT) and the learning of the target linguistic form (i.e., the results of the OEIT and the WPDT) are examined. As shown in the results of both grammar developmental tests in the previous section, a series of mixed-designed ANOVAs indicated significant Time effects but without Group and Group x Time interaction. Hence, this section examines the associations between the eye-tracking results (both the gain FPRT and the gain RRT) and all the participants' grammar learning gains (gain scores from the pretest to the delayed posttest on the OEIT and the WPDT) without including Group and its interaction with

each eye-tracking measure into the models.

First, the assumptions of multiple regression (i.e., absence of outliers and collinearity, linearity, homoscedasticity, and normality of residuals) were checked for each model. As shown in the correlation matrix presented in Table 4.31, the results of the gains scores from the pretest to the immediate posttest and the ones from the pretest to the delayed posttests were highly correlated with each other. Hence, only the gain scores from the pretest to the delayed posttest for each grammar test (the OEIT and the WPDT) were used as an outcome variable since learners' delayed posttest performances can be considered as an indication of “stable L2 development” (Issa & Morgan-Short, 2019, p. 400).

Table 4.31. A table of correlations among variables (The eye-tracking and the grammar test results)

	<i>FPRT</i> <i>_G1</i>	<i>RRT</i> <i>_G1</i>	<i>FPRT</i> <i>_G4</i>	<i>RRT</i> <i>_G4</i>	<i>OEIT</i> <i>_G1</i>	<i>OEIT</i> <i>_G2</i>	<i>WPDT</i> <i>_G1</i>	<i>WPDT</i> <i>_G2</i>
<i>FPRT_G1</i>	1	.17	.14	.14	.06	-.11	.07	.003
<i>RRT_G1</i>		1	.17	.06	.04	.02	.22*	.24*
<i>FPRT_G4</i>			1	.05	.15	.05	.23*	.06
<i>RRT_G4</i>				1	.08	-.05	-.01	-.07
<i>OEIT_G1</i>					1	.73***	.31**	.38***
<i>OEIT_G2</i>						1	.22*	.42***
<i>WPDT_G1</i>							1	.68***
<i>WPDT_G2</i>								1

Note. *ns* = not significant ($p > .05$), * $p < .05$, ** $p < .01$, *** $p < .001$

FPRT_G1: the gain *FPRT* for Text 1; *RRT_G1*: the gain *RRT* for Text 1

FPRT_G4: the gain *FPRT* for Text 4; *RRT_G4*: the gain *RRT* for Text 4

OEIT_G1: the gain scores from the pretest to the immediate posttest on the *OEIT*

OEIT_G2: the gain scores from the pretest to the delayed posttest on the *OEIT*

WPDT_G1: the gain scores from the pretest to the immediate posttest on the *WPDT*

WPDT_G2: the gain scores from the pretest to the delayed posttest on the *WPDT*

A hierarchical multiple regression was performed with the gain score of the *OEIT* as the outcome variable and each eye-tracking measures (i.e., the gain *FPRT* and *RRT* for Text 1, the gain *FPRT* and *RRT* for Text 4) as the predictor variables. Since the *RRT* is theoretically considered as the representation of L2 learners' effortful and non-automatic processing of

linguistic elements while reading L2 sentences and also since the eye-tracking results for Text 4 may have been influenced by their repeated engagements of the treatment tasks rather than purely showing their processing after engaging in the respective treatment task (see the overall research design in Chapter 3), the gain RRT for Text 1 was entered first into this hierarchical regression model. In the second step, the other predictor variables were entered into the model and then were assessed how much these variables contributed to the overall model additionally.

As shown in Table 4.32, the results of these two steps were summarized, showing very small amount of the variance of the OEIT gains were predicted by the gain RRT for Text 1 ($R^2 = 0.0001$, $p = .09$). Even after adding the other predictor variables, none of the predictor variables were predictive of the OEIT gains. As for the WPDT gains, somewhat different tendencies were indicated (see Table 4.33). The gain RRT for Text 1 significantly predicted the gain score on the WPDT, accounting for 6% of the variance ($R^2 = 0.06$, $p = .03$). By including the other predictor variables, the model prediction improved just by 1% but the model ($R^2 = 0.07$, $p = .23$). Based on these results of the separate hierarchical multiple regression, the gain RRT for the first passage (Text 1) was found to be predictive of the WPDT gains but the amount of the variance accounted for by the variable was small.

Table 4.32. Results of hierarchical multiple regression for variables (eye-tracking measures) predicting the gain scores from the pretest to the delayed posttest on the OPDT

	R^2	ΔR^2	b	95% CIs	SE	β	p
Step 1	0.0001						.09
Constant			14.41		1.77		< .001
RRT_G1			0.0001	[-.003, .004]	0.0017	0.01	.91
Step 2	0.02	0.02					<.83
Constant			14.31				< .001
RRT_G1			0.0004	[-.003, .004]	0.002	0.02	.84
FPRT_G1			-0.01	[-.02, .01]	0.01	-0.12	.32
RRT_G4			-0.001	[-.01, .004]	0.003	-0.04	.73
FPRT_G4			0.003	[-.01., .01]	0.01	0.07	.57

Notes.

FPRT_G1: the gain FPRT for Text 1; RRT_G1: the gain RRT for Text 1

FPRT_G4: the gain FPRT for Text 4; RRT_G4: the gain RRT for Text 4

OEIT_G1: the gain scores from the pretest to the immediate posttest on the OEIT

OEIT_G2: the gain scores from the pretest to the delayed posttest on the OEIT

WPDT_G1: the gain scores from the pretest to the immediate posttest on the WPDT

WPDT_G2: the gain scores from the pretest to the delayed posttest on the WPDT

Table 4.33. Results of hierarchical multiple regression for variables (eye-tracking measures) predicting the gain scores from the pretest to the delayed posttest on the WPDT

	R^2	ΔR^2	b	95% CIs	SE	β	p
Step 1	0.06						.03
Constant			15.18		1.82		< .001
RRT_G1			0.01	[.001, .01]	0.003	0.24	.03
Step 2	0.07	0.01					.23
Constant			15.21				< .001
RRT_G1			0.01	[.001, .01]	0.002	0.25	.03
FPRT_G1			-0.002	[-.02, .01]	0.01	-0.03	.80
RRT_G4			-0.003	[-.01, .01]	0.004	-0.09	.44
FPRT_G4			0.002	[-.01, .02]	0.01	0.03	.81

Notes.

FPRT_G1: the gain FPRT for Text 1; RRT_G1: the gain RRT for Text 1

FPRT_G4: the gain FPRT for Text 4; RRT_G4: the gain RRT for Text 4

OEIT_G1: the gain scores from the pretest to the immediate posttest on the OEIT

OEIT_G2: the gain scores from the pretest to the delayed posttest on the OEIT

WPDT_G1: the gain scores from the pretest to the immediate posttest on the WPDT

WPDT_G2: the gain scores from the pretest to the delayed posttest on the WPDT

CHAPTER 5: DISCUSSIONS

This chapter provides detailed discussions on the results reported in the previous chapter. Key findings are discussed with reference to each research question and its hypothesis. These key findings and their empirical and theoretical significance are also discussed in relation to previous studies and theories of SLA/ISLA.

The primary goal of this study was to re-visit and further advance Swain's Output Hypothesis by investigating detailed mechanisms of the noticing-triggering function of output. In particular, whether and how producing L2 output could induce learner-noticing of a problematic grammatical form in subsequent input and contribute to the acquisition of the form have not been fully addressed in previous studies due to various methodological issues (e.g., the operationalization of the type of output, measuring issues of output-induced noticing, the lack of examining the impact of modality difference of output [i.e., oral or written output], and a heavy reliance on a product-oriented research approach). Therefore, the current study aimed to further examine these long-standing issues in previous output and noticing studies by combining process- and product-oriented approaches through eye-tracking and L2 grammar developmental measures.

Specifically, the current study attempted to answer the following five research questions: To what extent does producing output induce learners' noticing of the target linguistic form (i.e., the past counterfactual/hypothetical conditional) in subsequent input? (RQ1); How does the modality of output affect the output-induced noticing of the target form in the subsequent input differently? (RQ2); To what extent does producing output contribute to the learning of the target linguistic form? (RQ3); How does the modality of output affect the learning of the target

linguistic form? (RQ4); and Is the amount of output-induced noticing associated with the overall learning of the target linguistic form? (RQ5).

In the following sections, first, whether and how L2 output and its different modalities triggered learner noticing (i.e., the noticing-triggering function of output) are discussed to address RQs1 and 2. Then, the impact of engaging in L2 output on the learning of the target grammatical form is discussed (for RQs3 and 4). Finally, the relationship between learner noticing induced by producing L2 output and the overall learning of the target grammatical form is discussed based on the findings from the correlation and regression analyses (RQ5).

Roles of L2 Output and Output Modality in Triggering Learner Noticing

The first research question (RQ1) aimed to test and further examine the mechanisms of the noticing-triggering function of Swain's Output Hypothesis with an online, objective measure (i.e., eye-tracking) by asking whether and how producing L2 output could trigger learner noticing in the subsequent input processing. In relation to the first question, the second research question (RQ2) investigated whether and how the modality of output (oral or written output) could influence the extent of output-induced noticing differently.

The overall eye-tracking results generally showed the opposite tendencies between the Input Only group and both output groups (i.e., the Oral Output and the Written Output groups), indicating that both output groups generally spent more time processing the areas of interest (AOIs) (i.e., the features of the target grammatical form) than did the Input Only group (see the Eye-Tracking section in Chapter 3 for detailed explanations and examples of the AOIs in the reading texts). These tendencies were more evident in the results of the RRT than the ones of the FPRT. Table 5.1 summarizes the gain fixation durations from the first reading (Reading 1) to the second reading (Reading 2) for both early and late measures of eye-tracking for both reading

texts (Text 1 and 4). As for the between-group differences indicated in the results of Text1, the Written Output group demonstrated significantly higher FPRT gains than did the other two groups, and both output groups (Written and Oral Output groups) showed significantly higher RRT gains than the Input Only group. Since the results of each of these early and late measures indicated a significant Group x Time interaction, these group differences suggested different tendencies between the groups on how learners processed the target linguistic features during the subsequent input after engaging in each respective treatment task (i.e., aural input [listening] or oral or written output [text-reconstruction]).

Table 5.1. *Summary of the gains from the first reading to the second reading*

<i>Eye-Tracking</i>	<i>Input Only</i>	<i>Oral Output</i>	<i>Written Output</i>
Text 1			
FPRT	- Gain * (-0.43)	- Gain ^{ns} (-0.26)	+ Gain * (0.39)
RRT	- Gain * (-0.63)	+ Gain * (0.60)	+ Gain * (0.60)
Text 4			
FPRT	- Gain ^{ns} (-0.40)	+ Gain ^{ns} (0.08)	+ Gain ^{ns} (0.11)
RRT	- Gain * (-0.53)	+ Gain ^{ns} (0.02)	+ Gain * (0.42)

Notes.

- Gain = a decreased gain from the first reading to the second reading
- + Gain = an increased gain from the first reading to the second reading
- * = significant gain; ^{ns} = negative gain
- The significant gains for each measure are bolded.
- Within-group effect sizes (Cohen's *d*) for the gain are presented in the brackets.

On the basis of these results, it can be suggested that engaging in written output induced the highest degrees of learner noticing in the subsequent input, which were evidenced by their significantly increased fixation durations on both early and late measures. Likewise, the Oral Output group also demonstrated significantly increased fixation durations but only on the late measure. In contrast, just engaging in input without having any opportunities to produce output did not seem to facilitate learner noticing, which was evidenced by their decreased fixation durations for processing the target linguistic form during the second reading. Therefore, the

hypotheses for the first and second research questions, which were hypothesized to support the noticing-triggering function of Swain's Output Hypothesis and the advantageous effects of the written modality of output were confirmed. Although the decreased FPRT demonstrated by the Oral Output group was contrary to the initial expectations, detailed explanations and discussions on these key findings and issues are discussed in the following sections.

First of all, the biggest and long-standing limitation of previous output and noticing studies was the lack of access to L2 learners' ongoing internal processes of noticing due to their reliance on the use of offline, indirect subjective measures, such as underlining, note-taking, retrospective questionnaire, and stimulated recalls. Therefore, these contrasting tendencies of the eye-tracking results between the output groups and the non-output (Input Only) group highlighted supportive evidence for the noticing-triggering function of output with fine-grained online objective data of L2 learners' cognitive processes. Particularly, these results empirically demonstrated how the "feedback loop to comprehended input" could function to induce learner noticing after producing output within the integrated model of L2 acquisition (Gass, 1997, p. 7) (also see Loew, 2015). These demonstrations of the detailed mechanisms of the noticing-triggering function of output have never been shown with online objective data in previous noticing and output studies and thereby were the biggest contribution of the current study to the theoretical development of Swain's (1985, 1995, 1998, 2005) Output Hypothesis, even though this theoretical model has been widely accepted and used as one of the crucial cognitive rationales for engaging in L2 output in the field of SLA/ISLA for years. Particularly, the current study clearly operationalized the output-induced noticing that the current study focused on (i.e., *noticing a form-meaning-function relationship* and *noticing the gap between IL and TL*) using Izumi's (2013) model of the four different types of noticing. This detailed operationalization of

noticing also contributed to a more fine-grained understanding of the type of learner noticing that could be induced through producing output (see Figure 2.2).

Although these eye-tracking results of the current study showed how engaging in L2 output could function as an internal priming device and induce learner noticing of their problematic form in subsequent input processing, the output-induced noticing that was focused on in this study was only one type of output-induced noticing that can be facilitated when learners receive relevant linguistic information as a form of subsequent input after producing output. Other types of output-induced noticing that could occur at different timings (e.g., while producing output) were not the focus of the current study. Thus, future studies that further examine how other types of output-induced noticing can be promoted by producing output at different stages of L2 acquisition processes will provide a more accurate and comprehensive understanding of the entire cognitive mechanisms of the noticing function of output.

As for Research Question 2, which investigated the impact of the modality difference of output (oral or written output) on learner noticing, both modalities of output similarly induced learner noticing of the target grammatical form as indicated by the group differences between both output groups and the Input Only group as well as the two output groups' significant gains from the first reading to the second reading on the results of the late measure (i.e., RRT). A different tendency was indicated between the two output groups only on the results of the early measure (i.e., FPRT) for Text 1, indicating a significantly higher FPRT gain for the Written Output group. As these results of the early and late measures of eye-fixation durations showed different tendencies, the initial hypothesis regarding the modality difference, particularly regarding the advantageous effects of a written modality of output on inducing a deeper level of learner noticing, was partially confirmed. However, the group difference in the results of the

FPRT between the two output groups highlighted much more fine-grained accounts of the noticing-triggering function of output at the different stages and levels of processes depending on the output modality difference.

While the late measure (RRT) indexes learners' effortful and non-automatic processing, such as controlled monitoring, the early measure (FPRT) represents learners' initial stage of sentence processing for comprehension (Clifton, Staub & Rayner, 2007; Conklin et al., 2018; Godfroid, 2020; Maie & Godfroid, 2022). Referring to Gass' (1988, 1997) integrated model of L2 acquisition from input to output (see the component of the comprehended input in the model depicted in Figure 2.1 in Chapter 2), the results of the FPRT are likely to represent the depth of processing at the stage of the comprehended input, which consists of two different levels of comprehension from relatively shallower semantic processing for mere comprehension to deeper detailed structural analyses of form-meaning-function mapping for further facilitation of grammatical knowledge development. Based on the current results of the FPRT, the group difference between the two output groups may have been explained by their different levels of initial processing of the target linguistic form to engage in deeper form-meaning-function mapping within the stage of the comprehended input, suggesting that the written output pushed learners to spend more time even at the initial stage of sentence processing for comprehension than did the oral output (also than did the Input Only group). Therefore, the Written Output group's increased FPRT in their second reading may be evidence of their deeper processing at the stage of the comprehended input depicted in Gass' (1988, 1997) integrated model of L2 acquisition.

Foreknowledge-Providing Function of Output

One unexpected but theoretically interesting finding was the non-significant gains for Text 4, which was demonstrated by both output groups on the FPRT and by the Oral Output group on the RRT (see Tables 4.9, 4.10, 4.12, 4.13, 4.18, and 5.1; Figures 4.3 and 4.4). This tendency was much more evident for the Oral Output group than the Written Output group because the Oral Output group did not show any significant eye-fixation gains from the first reading to the second reading on both FPRT and RRT. In contrast, the Input Only group did not show much difference in their eye-tracking results between Text 1 and Text 4, indicating relatively similar fixation duration decreases from the first reading to the second reading on Text 4 as they did on Text 1. It is not possible to rigidly compare the eye-tracking results of Text 1 with the ones of Text 4 due to the different word lengths included within the AOIs in each reading text. However, carefully examining both output groups' eye-tracking results of their first reading on Text 4 revealed that both output groups, especially the Oral Output group, spent substantially more time processing the target linguistic form during their first reading, which seemed to contribute to their smaller non-significant eye-fixation gains on the FPRT and the RRT.

First, the modality difference on the RRT can be explained by the findings of Zalbidea (2020), which reported that engaging in oral output led to the learners' higher rates of task-induced stress and anxiety than engaging in written output because the oral output modality requires learners to produce output spontaneously with less-processing time and thereby increasing the demand on their working memory load, which may have resulted in the modality difference on the results of the current study (also see Williams' inherent features of written production and their beneficial effects on grammar learning depicted in Figure 2.4 in Chapter 2).

More importantly, similar results of higher noticing from the first reading were reported by Song and Suh (2008), which also showed non-significant group differences between the two output groups (i.e., the written text-reconstruction and the picture-cued essay writing groups) and non-output group on their noticing (underlining) gains from the preceding input to the subsequent input even though both output groups showed significantly higher rates of underlining than that of the non-output group. Song and Suh (2008) argued that the preceding task instruction for the output group seemed to push the output group participants to focus on the target linguistic form even before working on the following output task because they knew from the preceding input phase that they needed to engage in the respective output tasks right after reading the text. That kind of foreknowledge of the following output task resulted in the output groups' greater attention to the target linguistic form from the first reading even before engaging in the output task. The same foreknowledge effect of output on learner noticing was reported by Yoshimura (2006), which indicated that providing foreknowledge of the following output task pushed learners to pay more attention to problematic linguistic forms for learners even before actually producing output.

The current study used the four reading texts during the treatment sessions and repeated the same instructional treatment procedures (First reading [Input 1] → Output or Aural Input [Task] → Second reading [Input 2]) for each reading text. Therefore, at the point of working on Text 4, all the learners in both output groups knew that they needed to work on the subsequent output task when they read the fourth text for the first time during the preceding input phase. Hence, the output groups' higher eye-fixation durations from the first reading provided supportive evidence for the findings of Song and Suh (2008) and Yoshimura (2006) and clearly demonstrated how having foreknowledge of the subsequent output task could induce learner

noticing even before producing output. On the basis of the findings from these previous output and noticing studies (e.g., Song & Suh, 2008; Yoshimura, 2006) and the eye-tracking results of the current study on Text 4, an additional function of output, named *the foreknowledge-providing function*, can be proposed as the fifth function of output in addition to the currently existing four functions (i.e., *the noticing-triggering function*, *the hypothesis testing function*, *the metalinguistic function*, and *the automatization function*). Although the current findings regarding the foreknowledge-providing function of output were theoretically and pedagogically very implicational, the current study was not specifically aimed to investigate the foreknowledge-providing function of output by comparing the results of the eye-tracking between Text 1 and Text 4 due to the different lengths of the AOIs between the two texts. Based on this methodological limitation, it is necessary to further examine the potential impact of providing foreknowledge about the subsequent output task on learner noticing using eye-tracking with a more suitable and rigid research design in future studies.

Roles of L2 Output and Output Modality in L2 Grammar Learning

Research questions 3 and 4 addressed whether engaging in L2 output contributes to the learning of the target grammatical form. To answer these questions, two different types of grammar developmental tests were used (i.e., the oral elicited imitation test [OEIT] and the written picture description test [WPDT]). The OEIT was conducted to measure the development of the learners' spontaneous receptive and productive knowledge using a test that required the learners to use the target linguistic knowledge in a different context from the one of the treatment task. Along with measuring the learners' generalizable knowledge development through the OEIT, the WPDT was also conducted to measure to what extent the learners can develop their control over the target linguistic form after engaging in each respective task.

Contrary to the original hypotheses, the results of the mixed-design ANOVA on the OEIT results indicated a significant main effect only for Time but not for Group or Group x Time interaction. In other words, all three instructional groups equally improved their test performances from the pretest to both posttests with large effect sizes regardless of the different instructional conditions. As for the results of the WPDT, somewhat different tendencies were observed. As shown on the WPDT results, the Written Output group demonstrated the highest gains on both posttests with large effect sizes, and the other two groups (the Input Only and the Oral Output groups) also significantly improved their test performances in the posttests but with smaller effect sizes. Despite these group differences in the effect sizes, the ANOVA results also indicated a significant main effect only for Time but not for Group or Group x Time interaction. Therefore, statistically convincing arguments about the group differences could not also be made on the results of the WPDT. Based on these results, the hypotheses for Research Questions 3 and 4, which predicted greater learning gains for both output groups, particularly for the written output group with its potential cognitive advantages of the written modality, were not confirmed with the results of both OEIT and WPDT.

One important question that needs to be discussed here is why both grammar developmental tests (i.e., the OEIT and the WPDT) failed to show statistically significant group differences even though the eye-tracking results clearly demonstrated higher degrees of output-induced noticing for both output groups. First of all, the lack of group differences in the OEIT results was due to the equal gains achieved by the Input Only group even though they just received and processed aural input without engaging in the text-reconstruct output task during the treatment sessions. One possible explanation for the Input Only group's significant gains on the OEIT is that just processing target linguistic exemplars multiple times in meaningful contexts

(four exemplar sentences in each of the four reading texts) through the written and aural input may have been beneficial to improve their OEIT performances (Izumi & Izumi, 2004). In addition, the picture cures for the text-reconstruction task were also provided to the Input Only group while they were listening to the aural input (see Figure 3.3 for sample picture cures in Chapter 3). Thus, these picture cures may have functioned as additional cognitive support to direct their attention to the target form along with processing the written and aural input during the treatment sessions.

However, the eye-tracking results showed opposite patterns between the Input Only group and the two output groups, indicating significantly less eye-fixation duration on both eye-tracking measures to process the AOIs of the target linguistic form by the Input Only group. Considering the lack of noticing demonstrated by the Input Only group, it was unlikely that the learners in the Input Only group engaged in “major cognitive processes in SLA, including noticing, hypothesis formulation and testing, conscious reflection, and automatization (Muranoi, 2007a, p. 76). Since noticing is the initial step but is also a crucial cognitive prerequisite for the long processes of L2 acquisition as depicted in Gass’ integrated model of L2 acquisition (Izumi, 2013; Schmidt, 1990, 2001, 2012), a more plausible explanation, especially for the results of the OEIT, is that the equal gains attained by the Input Only group were influenced by the test-practice effect of the OEIT. Previous ISLA studies that used the OEIT (e.g., Broszkiewicz, 2011; Ellis et al., 2006; Erlam et al., 2009; Li et al., 2016) also reported similar test-practice effect of the OEIT, which was demonstrated by a test-only control group who did not receive any instructional treatment but demonstrated steady improvements on their OEIT performances from the pretest to the immediate posttest and particularly from the pretest to the delayed posttest (see Suga & Loewen, 2020). Based on Suzuki and Koizumi’s (2020) descriptions, it was possible that

the learners of the current study became familiar with the test itself and then performed better for the second time and even better for the third time in the delayed posttest. Particularly, the participants of the current study were B2-level international undergraduate and graduate students who were studying at a US university. Thus, they possessed relatively high levels of functional English proficiency but still did not have full control over the target linguistic form at the point of the pretests. During the OEIT, test-takers of the OEIT were required to (1) listen to an exemplar sentence, (2) judge the content of the sentence by choosing true, false, or not sure based on their comprehension, and then (3) repeat the exemplar sentence accurately (see the procedures for the OEIT in the Methods section in Chapter 3). With an accurate preceding exemplar sentence that was provided to repeat (or reconstruct) each test item, it was possible that the participants of this study including the Input Only group started to be able to just repeat the test statements accurately without having much difficulty as they engaged in the OEIT repeatedly. Thus, their relatively high English proficiency may have enabled them to repeat each exemplar sentence accurately. However, it was difficult to identify which explanation was the case for the current results of the OEIT (i.e., whether the beneficial effects of input alone, the test-practice effects of the OEIT, or the combination of both contributed to the equal significant OEIT gains), because the current study did not include any test-only control group in the overall experimental design. If the current study had included a test-only control group, which only took the tests three times without receiving any instructional treatments, how much the input-only condition contributed to the learning gains in the posttests could have been shown. To address this methodological weakness, it is valuable to address this issue by including a test-only control group along with the replacement of the grammar developmental test with the type of test that does not provide accurate exemplars to learners just for the purpose of testing (e.g., an oral

picture description test) rather than the OEIT, especially when relatively high proficiency L2 learners are the targeted participants of the study.

As for the results of the WPDT, somewhat different tendencies were shown between the groups. The Written Output group attained the highest descriptive gains with large effect sizes, whereas the Input Only group and the Oral Output group showed smaller gains with small-medium effect sizes based on Plonsky and Oswald's (2014) guidelines for interpreting effect sizes in applied linguistics studies. As shown in Figure 4.7, the boxplots for the results of the WPDT also illustrated somewhat different tendencies between the Written Output group and the other two groups in terms of the Written Output group's changes of the middle 50 percent of the data and the median scores across the three testing sessions. However, these group differences were not statistically significant as indicated by the lack of significant Group and Group x Time interaction on the ANOVA results. Therefore, it is not possible to make statistically convincing arguments about these group differences and the beneficial effects of written output on the WPDT based on the ANOVA results. At the same time, the following correlation and regression analyses indicated associations between increased learner noticing (especially the increase in RRT) and the learners' WPDT gains. Thus, the slight group differences observed here are discussed in relation to the results of the correlation and the multiple regressions in the following section.

Relationship between Noticing and Output in SLA

Research Question 5 investigated the relationship between learner noticing that was gauged through the eye-fixation duration gains on the FPRT and the RRT and the learning of the target linguistic form (i.e., the gain scores on the OEIT and the WPDT). The correlation and the regression analyses indicated significant associations between the learners' increased RRT on

Text 1 and their WPDT gains from the pretest to the delayed posttest (see Tables 4.31 and 4.33), whereas the FPRT was not significantly predictive of the learners' grammar learning gains. The results of the OEIT were not associated with any of the eye-tracking results. Based on these results, the original hypothesis was not fully confirmed but the unexpected results suggested more detailed implications on the relationships between learner noticing, L2 output, and L2 grammar learning.

First of all, it was interesting that the FPRT gains were not significantly associated with any of the grammar learning gains whereas the RRT gains significantly predicted the WPDT gains. These contrasting results may have been attributed to the inherent nature of each type of learner noticing measured through the early and the late measures of eye-tracking. As discussed above, the early measure (FPRT) and the late measure (RRT) represent different stages of sentence processing (Conklin et al., 2018; Godfroid, 2020; Maie & Godfroid, 2022). Based on the results of the correlation and the regression analyses, one major finding regarding the relationship between learner noticing, L2 output, and the overall acquisition of L2 grammar was that the type of learner noticing which can be indexed by the late measure of eye-tracking (i.e., RRT) is important in L2 grammar development even though the RRT gains on Text 1 accounted for small amount of variance of the WPDT gains ($R^2 = 0.06$, $p = .03$). In other words, these results suggested that the deeper levels of learner noticing that were exhibited by L2 learners' non-automatic and controlled processing (i.e., re-reading) of the target linguistic features were necessary for the learning of the target linguistic form, but the initial stage of sentence processing that represented a relatively shallower semantic processing for comprehension (i.e., first-pass reading) may not have been sufficient for the development of L2 grammar knowledge.

These differential regression results between the early and the late measures led to another important question that needs to be discussed in relation to the slight group differences indicated above on the results of the WPDT. The eye-tracking results clearly demonstrated the noticing-triggering function of output, particularly for the written modality of output, indicating the highest degrees of learner noticing in the subsequent input for the Written Output group. Also, the regression results revealed a significant association between the increased RRT and the WPDT gains. Based on these results, it is crucial to discuss why the results of the WPDT only showed slight group differences without indicating any statistically significant between-group differences even though eye-tracking results demonstrated output-induced noticing and the regression results indicated the associations between the RRT gains and the WPDT gains.

These contrasting results may have been attributed to the small amount of variance accounted for by the RRT gains. Although the result was contrary to the initial expectation, it was not completely unpredictable considering the number and length of the treatment sessions and the amount of L2 practice that the learners engaged in during each respective treatment task. Furthermore, all the instructional treatment tasks were conducted relatively implicitly from the perspective of the instructor. Similar results were reported by Winke (2013), which was not exactly an output and noticing study but reported increased learner noticing (i.e., increased RRT) but the increased RRT was not reflected in the results of the grammar learning. For these contrasting results, Winke (2013) claimed, “the increase in the amount of noticing was not enough for immediately measurable acquisition to occur” (Winke, 2013, p. 341). As Schmidt (2001) stated that noticing is “the first step in language building” in the long process of L2 acquisition (p. 31), one important implication based on the lack of significant group differences especially in the WPDT results and the smaller amount of variance in the WPDT gains

accounted for by the increased RRT, may be that L2 learners do not necessarily and automatically incorporate all the linguistic features that they paid attention to or what they noticed into their interlanguage system. Throughout the entire long processes of L2 acquisition as depicted in Gass' (1988, 1997) integrated model of L2 acquisition, various linguistic information that learners noticed can be filtered out as it goes through various stages of cognitive processes of L2 acquisition from input to output.

Concerning the amount of output practice, a close re-examination of the previous output studies that reported beneficial effects of L2 output on grammar knowledge development (e.g., Ghari & Moinzadeh, 2011; Izumi, 2002; Kang, 2010; Li & He, 2017; Muranoi, 2007b; Russell, 2014; Shin, 2011; Song & Suh, 2008; Uggen, 2012), it turned out that these previous studies implemented two trials of the same output practice during each instructional treatment session following the sequence of Input → Output → Input → Output. Since the current study did not provide a second trial of output practice for each reading text, adding another trial of output practice for each reading text may have contributed to more solid grammar learning gains. However, this is another empirical question. Thus, the impact of the amount of output practice and the length of treatment sessions on learner noticing and the development of L2 grammar knowledge needs to be further examined in future studies.

Regardless of the limited effects of learner noticing on the overall L2 grammar learning as well as the smaller amount of associations between learner noticing and the WPDT gains, the findings of the current study still showed a significant association between the increased RRT and the L2 grammar development, which highlighted the detailed mechanisms of the noticing function of output and the roles of different types of learner noticing within the overall processes of L2 grammar learning depicted in Gass' (1988, 1997) integrated model of L2 acquisition. At

the same time, these results also highlighted that inducing learner noticing through L2 output may not necessarily be sufficient for leading to measurable differential grammar learning attainment.

CHAPTER 6: CONCLUSION AND LIMITATIONS

This chapter presents a summary of the key findings and their theoretical significance in the field of SLA/ISLA. Based on these findings, pedagogical implications for output-based instruction in L2 classrooms are discussed. At the end of this chapter, potential limitations and directions for future research are discussed.

Summary of the Key Findings of the Study

The present study aimed to re-visit and further advance the understanding of the noticing function of Swain's Output Hypothesis and address one of the long-standing major issues of previous output and noticing studies in the field of SLA/ISLA through a hybrid design of both process- and product-oriented research approaches. Specifically, this study was one of the first steps to shed light on whether and how producing L2 could induce learner-noticing of a problematic grammatical form in subsequent input and contribute to the acquisition of the form by employing two different levels of eye-tracking measures (i.e., the early and the late measures) as a sensitive, online objective measure of noticing and two grammar tests (i.e., the OEIT and the WPDT) as L2 developmental measures.

Overall, the results of the present study revealed the detailed mechanisms of the noticing-triggering function of output, the impact of differential output modalities (i.e., oral or written output), and the relationships between learner noticing and eventual L2 grammar learning attainment. Particularly, the eye-tracking results clearly demonstrated the opposite patterns between the two output groups (the Oral Output, and the Written Output groups) and the group that did not engage in any output (the Input Only group) during the instructional treatment, indicating significantly increased eye-fixation duration on the target grammatical features for both output groups and significantly decreased eye-fixation duration for the Input Only group.

Depending on the levels of the noticing measures (the early or the late measures of eye-tracking), the modality of L2 output influenced the degree of output-induced learner noticing differently, indicating a deeper level of learner noticing for the written output modality, which was evidenced by the results of both early and late measures (i.e., the gains on the FPRT and the RRT). Additionally, both output groups' increased eye-fixation duration even from the preceding input phase on Text 4 led to the proposition of an additional function termed the *foreknowledge-providing function* of output. Although these eye-tracking results clearly demonstrated how L2 output induced learner noticing in the subsequent input processing opportunities and slight group differences on descriptive results on the WPDT, the grammar test results did not show measurable beneficial effects of producing output on L2 grammar learning, which also highlighted the insufficiency of output-induced noticing itself for overall grammar acquisition.

All of these findings led to the conclusion that engaging in L2 output is important in inducing learner noticing as depicted as the “feedback loop” from output back to input in the integrated model of L2 acquisition (Gass, 1988, 1997; Leow, 2015). However, just engaging in producing output and processing subsequent input without accompanying any additional practice and instruction may not be sufficient for the immediate integration of problematic challenging grammatical structures.

Pedagogical Implications

In addition to these theoretical contributions, the findings of the present study suggested several pedagogical implications for L2 classroom instruction. As shown in the opposite patterns indicated on the results of the eye-tracking between the output groups and the Input Only group, engaging in output practice in the sequence of Input → Output → Input is a psycholinguistically valid instructional procedure in inducing learner noticing. Since learner noticing is an essential

cognitive prerequisite or foundation for L2 acquisition (Izumi, 2013; Schmidt, 1990, 1993, 1994, 1995, 2001, 2012; Schmidt & Frota, 1986), providing another input processing opportunity as subsequent input after producing L2 output creates an additional learning opportunity for learners rather than finishing L2 classroom lesson right after conducting L2 output practice. As Izumi (2013) described, two different types of noticing (i.e., *noticing a form-meaning-function relationship* and *noticing the gap between interlanguage and ideal target language uses*) are triggered in such a subsequent input processing opportunity. Therefore, it may be valuable to have classroom learners engage in reading or listening to related input after output following the instructional sequence of Input → Output → Input. At the same time, as shown in the results of the grammar developmental test, output-induced noticing may not necessarily lead to immediate grammar knowledge development. Therefore, providing additional L2 instruction together with the current study's sequence of Input → Output → Input may be more beneficial for grammar learning. One way of doing this is adding another round of output practice after the subsequent input as previous output studies indicated (e.g., Ghari & Moinzadeh, 2011; Izumi, 2002; Kang, 2010; Li & He, 2017; Muranoi, 2007b; Russell, 2014; Shin, 2011; Song & Suh, 2008; Uggen, 2012).

Next, the results of the study indicated that both output groups spent substantially more time processing the target linguistic features (or AOIs) from the first reading of Text 4 than they did during the first reading of Text 1. In other words, they processed the target linguistic features in Text 4 more carefully even before engaging in the subsequent output practice. This was because, at the point of Text 4, they had already engaged in the output task three times following the same Input → Output → Input procedures for Texts 1, 2, and 3, and thus they knew that they needed to work on the following output task (oral or written) even from the point of their first

reading of Text 4. Based on this observed phenomenon and the findings from previous studies (e.g., Song & Suh, 2008; Yoshimura, 2006), the foreknowledge-providing function of output was proposed as the fifth function of output, which is having foreknowledge of the subsequent output task could induce learner noticing even before producing output. This additional function of output (i.e., the foreknowledge-providing function of output) highlights the importance of incorporating L2 output practice as a routine activity into an L2 teaching curriculum and/or a syllabus design. By so doing, learners are more likely to focus on their problematic linguistic forms and process these more carefully throughout the entire class hours even before engaging in L2 output practice, eventually providing them with more opportunities to initiate focus on form and incorporate their problematic linguistic forms into their developing interlanguage system.

The current study implemented output practice implicitly from the perspective of the instructor (the researcher) because eye-tracking is a very sensitive measure of learners' cognitive processes and it was crucial to eliminate all irrelevant variables that could potentially influence the learners' cognitive processes as empirical research. However, in L2 classroom contexts, L2 teachers do not need to (or should not) just focus on a limited number of certain instructional variables because the primary objective of L2 instruction is to maximize L2 learners' learning through their classroom instruction. Therefore, combining output practice with other various instructional techniques (e.g., explicit rule explanations) may be one way to make the instruction more beneficial (see Goo et al., 2015; Kang et al., 2019; Koyanagi, 2016; Norris & Ortega, 2000; Spada & Tomita, 2010). For example, previous studies (e.g., Li, Ellis, & Zhu, 2016; Muranoi, 2007b; Shintani, 2019) reported beneficial effects of combining text-reconstruction output practice with explicit grammar explanations. However, it is also another empirical question

whether adding additional instructional variables to the current treatment design can be more beneficial or not.

Finally, in relation to the relatively implicit implementation of the current instructional treatment sessions, the learners in both output groups directed their attention to the target linguistic form without explicitly introducing the target linguistic form during the instructional treatment. Regarding the output-induced noticing observed for both output groups, the learners identified their problematic linguistic features and then directed their attention to these features by themselves. In this study, the researcher intentionally designed the content of the instructional materials by eliminating any low-frequency words and late-acquired, difficult grammar features other than the targeted form so that the target form could be the only difficult linguistic feature for the participants of the study. However, if output activities can induce learner noticing based on each learner's respective linguistic needs, classroom L2 learners may engage in focus on form depending on their specific needs, proficiency levels, and specific linguistic challenges for each learner. For example, less proficient learners are likely to pay more attention to their immediate needs, such as lexical items and basic grammatical features while more proficient learners may focus on more advanced grammatical forms even within the same L2 classroom (e.g., Hanaoka, 2007; Hanaoka & Izumi, 2012; Leaser, 2008; Swain & Lapkin, 1995; Williams, 1999, Uggen, 2012). Therefore, the findings of the study suggested the potential of output practice as a way to enhance learner noticing, which can be initiated by the learners themselves and thereby may be an important instructional technique to accommodate various learners with varying linguistic proficiency in L2 classrooms.

Limitations and Directions for Future Research

Despite these theoretical and pedagogical contributions of the findings of the current study, it must be noted that this study was also limited by several methodological limitations that need to be considered in future research. First, as discussed in the discussion chapter, the major limitation of the study was the potential test-practice effect indicated based on all three groups' relatively equal learning gains. Particularly when the targeted participants are relatively high proficiency L2 learners as the ones in this study, the type of grammar developmental tests that provide accurate exposures of the targeted form(s) (or additional positive evidence) for the learners as did the OEIT should be avoided. For example, an oral picture description test may have been a good counterpart for the WPDT used in the current study.

Second, another limitation was also related to the design of the study. The current study operationalized output-induced noticing as the type of learner noticing that occurred in the subsequent input after producing output. However, the noticing-triggering function of output also includes the types of learner noticing that could be induced while learners are engaging in output practice (i.e., *noticing holes in one's interlanguage system* and *noticing the gap in one's ability*) (see Izumi, 2013; also see Figure 2.2 for different types of noticing at different timing of the L2 processing). In future studies, it is also valuable to investigate how L2 output induces learner noticing while learners are producing output and how the different types of output-induced noticing are related to the noticing-triggering function of output and also to the overall processes of L2 acquisition. For example, such while-processing output-induced noticing can be investigated using verbal reports, such as stimulated recalls with learners' production data or audio recordings as the stimulus or think-aloud, which can tap into learners' cognitive processes while producing output.

Third, although the current study also conducted the stimulated recalls during the data collection session as described in the overall research design (see Figure 3.1 and Table 3.2), the data from the retrospective subjective data was not included in this study. The eye-tracking results can show how much learners pay attention to certain aspects of the target linguistic features based on sensitive, online objective measures of their eye-movements (e.g., both early and late measures of their fixation duration in this study). As Godfroid et al. (2013) claimed, verbal reports, such as stimulated recall and think-aloud protocols, can better shed light on the aspects of learners' linguistic processing particularly their conscious awareness than eye-tracking, which is better for measuring the amount of attention and the locus of their attentional resource. Therefore, it is valuable to also incorporate verbal reports to triangulate the data with verbal reports (e.g., Wang & Pellicer-Sanchez, 2023). Particularly, the participants' processing data that was collected but was not included in this dissertation (i.e., the recordings of all participants' text-reconstruction performances and their subjective, retrospective accounts of their noticing [i.e., stimulated recall data]) may provide more detailed accounts for the slight differences reported in the WPDT results between the Written Output group and the other two groups.

Finally, the potential attrition rate of the participants was relatively high due to technical difficulties concerning the use of the eye-tracker, such as failed calibrations and validations, and failed recording, especially at the beginning of the data collection period. In addition to the participants who were removed from the analyses due to their vocabulary level, proficiency, and test-performances that were considered as outliers ($n = 10$), fourteen potential participants were not included in this study due to these technical reasons. It may be very helpful to create a system to share specific technical difficulties and troubleshooting techniques among SLA/ISLA

researchers who have used eye-tracking measures so that future researchers do not need to repeat the same technical difficulties that previous eye-tracking researchers went through.

Although the current study was limited by these methodological limitations, the current study revisited Swain's Output Hypothesis and further examined the detailed cognitive mechanisms of the noticing-triggering function of output through the use of a sensitive, online objective measure of learner noticing called eye-tracking. Particularly, the findings of the current study shed light on how engaging in L2 output could induce learner noticing in the subsequent input-processing opportunities. Future studies that will address the limitations of the current study will provide more detailed accounts of the noticing-triggering function of output and contribute to further understanding of how systematic manipulation of the mechanisms of L2 learning facilitate the development of L2 learners' interlanguage system.

REFERENCES

- Alsulami, S. Q. (2016). Testing the Noticing Function of the Output Hypothesis. *English Language Teaching*, 9(2), 136-141. doi:10.5539/elt.v9n2p136
- Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, 35(4), 689–725. <https://doi.org/10.1017/S0272263113000429>
- Boers, F. (2021). *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns*. Routledge.
- Broszkiewicz, A. (2011). The effect of focused communication tasks on instructed acquisition of English past counterfactual conditionals. *Studies in Second Language Learning and Teaching*, 1(3), 335-262. <https://doi.org/10.14746/ssllt.2011.1.3.3>
- Busterrechea, M., García Mayo, M. D. P., & Leeser (2014). Pushed output and noticing in a dictogloss: Task implementation in the CLIL classroom. *Porta Linguarum* 22, 7-22. <http://hdl.handle.net/10481/53643>
- Celce-Murcia, M., & Larsen-Freeman, D. (2016). *The grammar book: Form, meaning, and use for English language teachers*. Heinle Cengage Learning.
- Cho, M. (2018). Task complexity and modality: Exploring learners' experience from the perspective of flow. *The Modern Language Journal*, 102(1), 162-180. <https://doi.org/10.1111/modl.12460>
- Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*(pp. 341–371). Elsevier. <https://doi.org/10.1016/B978-008044980-7/50017-3>
- Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistic research*. Cambridge University Press.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), 31-51. <https://doi.org/10.1177/026553229000700104>
- de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46(3), 529-555. <https://doi.org/10.1111/j.1467-1770.1996.tb01246.x>
- de Graaff, R., & Housen, A. (2009). Investigating the Effects and Effectiveness of L2 Instruction. In M. Long & C. J. Doughty (Eds.), *The handbook of language teaching*, (pp. 726–755). Blackwell.
- Doughty, C. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition*, 13(4), 431–469. <https://doi.org/10.1017/S0272263100010287>

- Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 206-257). Cambridge University Press.
- Duolingo (2021). *Analysis of the scoring and reliability for the Duolingo English test*. Duolingo, Inc. <https://d23cwzsbkjb45.cloudfront.net/media/resources/standards/scoring.pdf>
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. In R. Ellis (Eds.), *Form-focused instruction and second language learning* (pp. 1-46). Blackwell Publishers.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141-172. <https://doi.org/10.1017/S0272263105050096>
- Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, 27(3), 431-463. <https://doi.org/10.1093/applin/aml022>
- Ellis, R. (2008). *The study to second language acquisition* (2nd ed.). Oxford University Press.
- Ellis, R., Loewen, S. D., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27, 464–491. <https://doi.org/10.1093/applin/aml001>
- Erlam, R., & Akakura, M. (2016). New developments in the use of elicited imitation. In A. Mackey & E. Marsden (Eds.), *Advancing Methodology and Practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 105-123). Routledge.
- Gass, S. (1988). Integrated research areas: A framework for second language studies. *Applied Linguistics*, 9(2), p. 198-217. <https://doi.org/10.1093/applin/9.2.198>
- Gass, S. (1997). *Input and interaction, and the second language learner*. Lawrence Erlbaum Associates.
- Gass, S. (2013). *Second language acquisition: An introductory course*. (4th ed.). Routledge.
- Gass, S., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research*. (2nd ed.). Routledge.
- Ghari, A., & Moinzadeh, A. (2011). The effects of output task types on noticing and learning of English past modals: A case of intermediate Persian adult learners of English. *Journal of Language Teaching and Research*, 2(5), 1180-1191. doi:10.4304/jltr.2.5.1180-1191

- Gilabert, R., Manchón, R., & Vasylets, L. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics*, 36, 117–135. <https://doi.org/10.1017/S0267190515000112>
- Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners' eye-tracking data. In R. Leow (Eds.), *The Routledge handbook of second language research in classroom learning* (pp. 44-57), Routledge.
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.
- Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Inside the learner's mind: Cognitive processing and second language acquisition* (pp. 169–197). John Benjamins Publish Company.
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483-517.
- Godfroid, A., & Spino, L. A. (2016). Under the radar: Triangulating think-alouds and finger-tracking to detect the unnoticed. In A. Mackey, & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 94-111). Taylor and Francis.
- Godfroid, A., & Uggen, M. S. (2013). Attention to irregular verbs by beginning learners of German: An eye-movement study. *Studies in Second Language Acquisition*, 35(2), 291-322. <https://doi.org/10.1017/S0272263112000897>
- Ha, H. T. (2021). Exploring the relationships between various dimensions of receptive vocabulary knowledge and L2 listening and reading comprehension. *Language Testing in Asia*, 11(1), 1-20. <https://doi.org/10.1186/s40468-021-00131-8>
- Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, 32(3), 465–491. <https://psycnet.apa.org/doi/10.1017/S0272263110000045>
- Hanaoka, O. (2006a). Exploring the role of models in promoting noticing in L2 writing. *JACET Bulletin*, 42, 1–13.
- Hanaoka, O. (2006b). Noticing from models and reformulations: A case study of two Japanese EFL learners. *Sophia Linguistica*, 54, 167–192.
- Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research*, 11(4), 459-479. <https://doi.org/10.1177/1362168807080963>

- Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing, 21*(4), 332-347. <https://doi.org/10.1016/j.jslw.2012.09.008>
- Hanaoka, O., & Izumi, S. (2021). Expanding research agendas: Directions for future research on attention and writing. In R. Manchón & C. Polio (Eds.), *The Routledge handbook of second and foreign language writing* (pp. 312-324). Routledge.
- Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing 9*(2), 122–159. <https://doi.org/10.1016/j.asw.2004.06.001>
- Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 23*(1), 403–430. DOI:10.26686/wgtn.12560354
- Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of second language writing, 11*(4), 329-350. [https://doi.org/10.1016/S1060-3743\(02\)00091-7](https://doi.org/10.1016/S1060-3743(02)00091-7)
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition 24*(4), 541-77. <https://doi.org/10.1017/S0272263102004023>
- Izumi, S. (2003). Comprehension and production processes in second language learning. *Applied Linguistics, 24*(2), 168-196. <https://doi.org/10.1093/applin/24.2.168>
- Izumi, S. (2013). Noticing and L2 Development: Theoretical, Empirical, and Pedagogical Issues. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 37-50). University of Hawai'i, National Foreign Language Resource Center.
- Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition?. *TESOL Quarterly, 34*(2), 239-278. <https://doi.org/10.2307/3587952>
- Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*(3), 421–452. <https://www.jstor.org/stable/44486913>
- Izumi, Y., & Izumi, S. (2004). Investigating the effects of oral output on the learning of relative clauses in English: Issues in the psycholinguistic requirements for effective output tasks. *Canadian Modern Language Review, 60*(5), 587–609. <https://doi.org/10.3138/cmlr.60.5.587>
- Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition, 39*(3), 401-430. <https://doi.org/10.1017/S027226311600019X>

- Issa, B. I., & Morgan-Short, K. (2019). Effects of external and internal attentional manipulations on second language grammar development: An eye-tracking study. *Studies in Second Language Acquisition*, 41(2), 389-417. <https://doi.org/10.1017/S027226311800013X>
- Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, 40(4), 755-780. <https://doi.org/10.1017/S0272263118000165>
- Kang, E. Y. (2010). Effects of output and note-taking on noticing and interlanguage development. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 10(2), 19-36. <https://doi.org/10.7916/salt.v10i2.1428>
- Kormos, J. (2014). *Speech production and second language acquisition*. Routledge.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Leeser, M. J. (2008). Pushed output, noticing, and development of past tense morphology in content-based instruction. *Canadian Modern Language Review*, 65(2), 195–220. <https://doi.org/10.3138/cmlr.65.2.195>
- Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, 22(4), 557–584. <https://www.jstor.org/stable/44486935>
- Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.
- Leow, R. P. (2019). Theoretical underpinnings and cognitive processes in instructed SLA. In R. Leow (Eds.), *The Routledge handbook of second language research in classroom learning* (pp. 15-27). Routledge.
- Leung, J. H., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, 33(1), 33-55. <https://doi.org/10.1017/S0272263110000525>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics*, 40(3), 693-722. <https://doi.org/10.1017/S0142716418000796>
- Li, W., & He, X., (2017, October). *The effectiveness of written output on promoting L2 learners' attention: An eye-tracking study*. Poster session presented at Second Language Research Forum 2017, Columbus, OH.

- Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han & E. Park (Eds.), *Understanding second language process* (pp. 24-44). Multilingual Matters.
- Lim, G. S., Geranpayeh, A., Khalifa, H., & Buckendahl, C. W. (2013). Standard setting to an international reference framework: Implications for theory and practice, *International Journal of Testing*, 13(1), 32-49. <https://doi.org/10.1080/15305058.2012.678526>
- Loewen, S. (2015). *Introduction to instructed second language acquisition*. Rutledge.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback?. *Studies in second language acquisition*, 22(4), 471-497. <https://psycnet.apa.org/doi/10.1017/S0272263100004022>
- Maclaughlin, B. (1990). *Restructuring*. *Applied Linguistics*, 11(2), 113-128. <https://doi.org/10.1093/applin/11.2.113>
- Maie, R., & Godfroid, A. (2022). Controlled and automatic processing in the acceptability judgment task: An eye-tracking study. *Language Learning*, 72(1), 158-197. <https://doi.org/10.1111/lang.12474>
- Manchón, R. M. (2011). *Learning-to-write and writing-to-learn in an additional language*. John Benjamins Publishing Company.
- Manchón, R. M. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp. 27–53). John Benjamins.
- Muraoka, Y. (2006). The effects of output and explicit metalinguistic explanation on the acquisition of English articles. *Educational Studies (International Christian University)*, 48, 217-226.
- Muranoi, H. (2007a). Output practice in the L2 classroom. In R. Dekeyser (Eds.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). Cambridge University Press.
- Muranoi, H. (2007b). Focus on form through guided summarizing and EFL learners' interlanguage development. *Journal of Institute for Research in English Language and Literature: Tohoku Gakuin University*, 33, 15-59. <https://tohoku-gakuin.repo.nii.ac.jp/records/24332>
- Muranoi, H. (2012, November). *Second language acquisition theory and English education (Daini genngo shutokuriron to eigokyoiku)*. Presented at Sophia University Kaken (Grants-in-Aid for Scientific Research) Lecture 2012, Tokyo, Japan.

- Nassaji, H. (2020). The importance of using multiple measures or data sources in L2 instructional research. *Language Teaching Research*, 24(2), 131-135. <https://doi.org/10.1177/1362168820906908>
- Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). *The association between TOEFL iBT® test scores and the common european framework of reference (CEFR) levels (Research Memorandum No. RM-15-06)*. Educational Testing Service. <https://www.ets.org/Media/Research/pdf/RM-15-06.pdf>
- Philip, J. (2013). Noticing hypothesis. In P. Robinson (Eds.), *The Routledge encyclopedia of second language acquisition* (pp. 464-466). Routledge.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. <https://doi.org/10.1111/lang.12079>
- Polio, C. (2020). Can writing facilitate the development of grammatical competence?: Advancing research agendas. In R. M. Manchón (Eds.), *Writing and language learning: Advancing research agenda* (pp. 381-401). Routledge.
- Polio, C. (2022). L2 writing and grammar development. In R. M. Manchón, & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 169-182). Routledge.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://psycnet.apa.org/doi/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Robinson, P. (1995). Attention, memory and the “noticing” hypothesis. *Language Learning*, 45(2), 283–331. <https://doi.org/10.1111/j.1467-1770.1995.tb00441.x>
- Russell, V. (2014). A closer look at the output hypothesis: The effect of pushed output on noticing and inductive learning of the Spanish future tense. *Foreign Language Annals*, 47(1), 25–47. <https://doi.org/10.1111/flan.12077>
- Shin, M. (2011). Effects of output tasks on Korean EFL learners' noticing and learning of English grammar. *Foreign Languages Education*, 18(2), 127-163.
- Schmidgall, J. (2021). *Mapping the redesigned TOEIC bridge ® test scores to proficiency levels of the common European framework of reference for languages. (Research Memorandum No. RM-21-01)*. Educational Testing Service.

- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, 13, 206-226. <https://doi.org/10.1017/S0267190500002476>
- Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Eds.), *Implicit and explicit learning of language* (pp. 165-209). Academic Press.
- Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Eds.), *Attention and awareness in foreign language learning* (pp. 1-63). University of Hawai'i Press.
- Schmidt, R. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 3-32). Cambridge Applied Linguistics.
- Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt, & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 27-50). De Gruyter Mouton.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner. In R. Day (Eds.), *Talking to learn: Conversation in second language acquisition* (pp. 237-369). Newbury House.
- Shintani, N. (2019) The roles of explicit instruction and guided practice in the proceduralization of a complex grammatical structure. In R. M. DeKeyser & G. P. Botana (Eds.). *Doing SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability* (pp. 83-106). John Benjamins Publishing Company.
- Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language learning*, 64(1), 103-131. <https://psycnet.apa.org/doi/10.1111/lang.12029>
- 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. <https://doi.org/10.1016/j.system.2007.09.006>
- Spada, N., & Lightbrown, P. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly*, 42(2), 181-207. <https://www.jstor.org/stable/40264447>
- Suh, B. R. (2010). *Written feedback in second language acquisition: Exploring the roles of type of feedback, linguistic target, awareness, and concurrent verbalization* (Unpublished doctoral dissertation). Georgetown University.
- Suga, K. & Loewen, S. (2023). Potential Test-Learning Effects of an Oral Elicited Imitation Test: Methodological Considerations for Form-Focused Instruction Studies. *Research Methods in Applied Linguistics*. 2(1), 100035. <https://doi.org/10.1016/j.rmal.2022.100035>

- Suzuki, Y. & Koizumi, R. (2020). Using equivalent test forms in SLA pretest-posttest design research. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 457-467). Routledge.
- Swain, M. (1985). Communicative competence: Some role of comprehensible input and comprehensible output in its development, In S. Gass & C. Medden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford University Press.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). Cambridge University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 471-481). Cambridge University Press.
- Swain, M. & Lapkin, S. (1995). Problems in output and cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 371-91.
<http://dx.doi.org/10.1093/applin/16.3.371>
- Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology, *TOEFL iBT Research Report*, ETS.
- Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183–203.
<https://doi.org/10.1017/S0272263100012870>
- Uggen, M. S. (2012). Reinvestigating the noticing function of output. *Language learning*, 62(2), 506-540. <https://doi.org/10.1111/j.1467-9922.2012.00693.x>
- VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex.
- Vasylets, O., & Gilabert, R. (2022). Task effects across modalities. In R. M. Manchón, & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 39-51). Routledge.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test. *ITL-International Journal of Applied Linguistics*, 168(1), 33-69.
<https://doi.org/10.1075/itl.168.1.02web>
- Williams, J. (1999). Learner-generated attention to form. *Language learning*, 49(4), 583-625.
<https://doi.org/10.1111/0023-8333.00103>

- Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of second language writing*, 21(4), 321-331. <https://doi.org/10.1016/j.jslw.2012.09.007>
- Williams, J. N. (2005). Learning without awareness. *Studies in second language acquisition*, 27(2), 269-304. <https://doi.org/10.1017/S0272263105050138>
- Winke, P. M. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, 35(2), 323-352. <https://doi.org/10.1017/S0272263112000903>
- Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33(4), 497-528. <https://doi.org/10.1177/0265532215594643>
- Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to differences in reading behaviour, text comprehension and noticing of language form?. *Language Teaching Research*, 10(4), 419-434. <https://doi.org/10.1191/1362168806lr204oa>
- Zalbidea, J. (2020). A mixed-methods approach to exploring the L2 learning potential of writing versus speaking. In R. M. Manchón (Eds.), *Writing and language learning: Advancing research agendas* (pp. 207-230). John Benjamins Publishing Company.
- Zalbidea, J. (2021). On the scope of output in SLA: Task modality, salience, L2 grammar noticing, and development. *Studies in Second Language Acquisition*, 43(1), 50-82. <https://doi.org/10.1017/S0272263120000261>
- Zalbidea, J., & Sanz, C. (2020). Does learner cognition count on modality? Working memory and L2 morphosyntactic achievement across oral and written tasks. *Applied Psycholinguistics*, 41(5), 1171-1196. <https://doi.org/10.1017/S0142716420000442>

APPENDIX A: WEBB, SASAO, AND OLIVER'S (2017) UPDATED VOCABULARY

LEVELS TEST

This is test that looks at how well you know useful English words. Put a check under the word that goes with each meaning. Here is an example.

	game	island	mouth	movie	song	yard
land with water all around it						
part of your body used for eating and talking						
piece of music						

It should be answered in the following way.

	game	island	mouth	movie	song	yard
land with water all around it		✓				
part of your body used for eating and talking			✓			
piece of music					✓	

1,000 Word Level

	boy	rent	report	size	station	thing
how big or small something is						
place buses and trains go to						
young man						

	ear	gold	lake	letter	office	people
information sent to people						
men and women						
place for working						

	fellow	hat	ice	joke	light	system
funny story						
man or boy						
something worn on your head						

date forest mistake news record shop

latest information
place with many trees
something that is not right

bar conversation neighbor rain rubbish shirt

person who lives nearby
things that are thrown away
type of clothing

continue cook phone pull sail share

hold and move something toward yourself
keep happening
use together with others

enter finish happen own sing worry

end
go inside
have something that is yours

arrive collect consider glance need pack

look quickly at something
reach the place you are going
think about something

affordable beautiful boring dry rough tall

higher than normal
not flat
not interesting

closed dirty empty musical orange sad

having nothing
not clean
unhappy

bar conversation neighbor rain rubbish shirt

person who lives nearby

things that are thrown away

type of clothing

continue cook phone pull sail share

hold and move something toward yourself

keep happening

use together with others

enter finish happen own sing worry

end

go inside

have something that is yours

arrive collect consider glance need pack

look quickly at something

reach the place you are going

think about something

affordable beautiful boring dry rough tall

higher than normal

not flat

not interesting

closed dirty empty musical orange sad

having nothing

not clean

unhappy

develop identify improve possess provide sew

give
have
make better

complain increase pray produce recognize whip

get larger
know and remember
make

curious defensive energetic nervous various wicked

different kinds of things
very bad
wanting to know

advanced cruel lone stiff typical upset

at a high level
not kind
single

3,000 Word Level

colleague fate fee hint status talent

ability or skill
clue
person you work with

circuit clinic format origin peak routine

place where you can see a doctor
top
what you usually do each day

agency heel pavement penalty principal youth

back of your foot

person in charge of a school

punishment

element jail joint objective portrait variety

goal

picture

place where criminals are kept

defeat infant nuclear outrage prospect rival

loss

person you oppose

small child

coincide derive devote permit publish regret

feel bad about doing something

give all your time and attention

happen at the same time

civilize discharge graduate imply merge perceive

join

release

suggest

assault bargain compete dedicate nominate restrain

attack

hold back

try to win

fundamental humorous interior numerous prompt religious

basic

many

on time

	legislative	mechanic	mortal	random	rear	reluctant
back of something						
can die						
without order						

4,000 Word Level

	auction	bullet	fever	flock	outlet	skull
group of birds						
high body temperature						
sale where people place bids						

	archive	ash	mat	moisture	physics	tile
place where old books are kept						
powder left after something burns						
science subject						

	pioneer	dictionary	immigration	petition	romance	thigh
book with information given for each word						
first person to do something						
paper that people sign						

	acid	cafe	deadline	deficiency	texture	thesis
lack						
place for buying and drinking coffee						
time limit						

	avenue	brass	departure	hood	hut	premier
cover for your head						
small house						
type of metal						

appall invade mutter refine roast unveil

cook over fire
enter by force
make pure

aspire exert gossip minimize poke postpone

make smaller
push with your finger
try to reach a goal

adhere fracture originate peel sparkle terminate

do what is expected
end
give off small flashes of light

amateur arrogant cognitive infinite judicial monetary

having no limits
not professional
overly proud

delicate dull miserable noble peculiar refreshing

breaks easily
unselfish and morally good
very unhappy

5,000 Word Level

calf epidemic foam landmark token trumpet

illness spread quickly that affects many
people
many bubbles
young cow

comb ivory pants rainbow vegetarian zip

containing no meat
hard white substance
tool for styling hair

analogy captive remainder renovation ribbon vest

comparison between two things
person kept somewhere unwillingly
what is left

butcher chalk grape ornament pier wallet

container for money
person who cuts and sells meat
place for boats to dock

ammunition crab dusk nucleus revenge spectator

beginning of night
center
person who watches

abolish apprehend chuckle erode replicate segregate

end
keep apart
slowly make smaller

duplicate emigrate hurl perch revolt swirl

copy
fight violently against
sit in a high place

amplify evaporate grunt mitigate recollect tow

disappear
make larger
remember

blunt	fabulous	horrified	numb	singular	volatile
--------------	-----------------	------------------	-------------	-----------------	-----------------

not sharp

without feeling

wonderful

brisk	extinct	fragrant	splendid	tolerant	trivial
--------------	----------------	-----------------	-----------------	-----------------	----------------

fast

having no living members

of little importance

APPENDIX B: RECRUITING FLYER

PARTICIPANTS NEEDED FOR LANGUAGE LEARNING RESEARCH \$45 for your participation

We are recruiting research participants for an English learning research which will explore how second language English speakers engage in English learning tasks and perform a series of tests over time.

What?

- Work on 4 sessions (3 hours in total)
- You need to come to a lab in Wells Hall 3 times.
- Your eye movements will be recorded while reading English texts.
- Sessions 2 & 3 need to be two consecutive days.
- The final reviewing session will be conducted online through Zoom.
 - Session 1 → 30 minutes (Wells Hall) → 4-7 days
 - Session 2 → 60 minutes (Wells Hall) → 1 day (2 consecutive days)
 - Session 3 → 60 minutes (Wells Hall) → 2 weeks
 - Session 4 → 30 minutes ← (Online reviewing session through Zoom)

How much?

- Receive \$45 after completing all the 4 sessions (3 hours in total)

Who?

- **International undergraduate & graduate students**
- **TOEFL iBT 72-95, IELTS 5.5-6.5, or Duolingo English Test 100-125, TOEIC around 700-900 (If you only have TOEIC scores, please contact me.)**
- **Stayed in the US less than 5 years (4 years or less)**
- Received primary and secondary education in your first language other than English (**NOT English**)

When?

- Choose 3 days for lab sessions (Wells Hall) + 1 day for an online session



If you are interested in participating in this study, **please fill out the interest form.** https://msu.co1.qualtrics.com/jfe/form/SV_9FeZXvoEYlz2Ssm

Or contact **Kiyo Suga** (sugakiyo@msu.edu).

Phone: 5172155326

Second Language Studies at Michigan State University



APPENDIX C: A PASSAGE USED FOR THE PRACTICE READING

The Practice Text

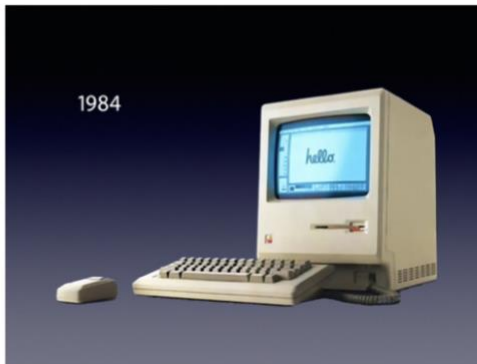
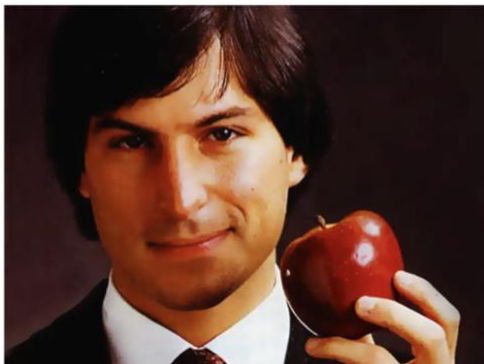
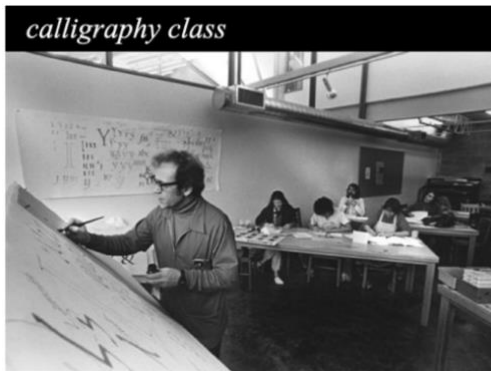
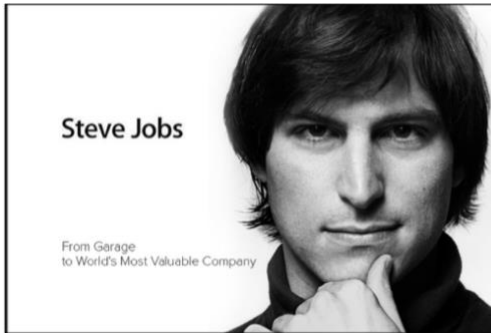
Michael Jackson, an American singer, was often called the “King of Pop.” His albums and videos sold amazingly well – more than seven hundred million copies! Michael was also a great dancer. He died in 2009 at the age of 50, but he is still popular around the world.

He started to sing professionally when he was only five. So, he never had time to enjoy his childhood. He thus came to believe that every child should have a good childhood.

APPENDIX D: SLIDES FOR THE ORAL INTRODUCTION FOR THE READING

TEXT

Text 1: Steve Jobs 1

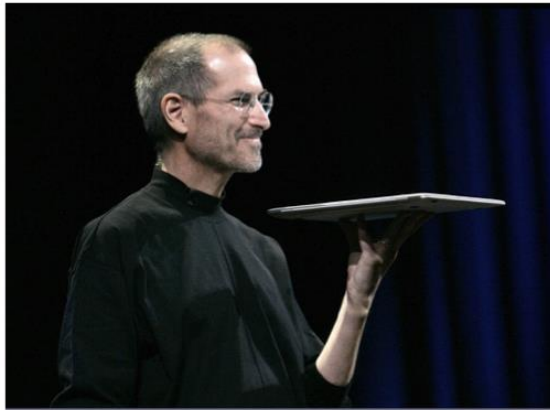


Text 2: Steve Jobs 2



1985







Text 3: Ichiro Suzuki



3,000 HITS CLUB

USA TODAY SPORTS

3,000 Ichiro Suzuki is the 30th player in history with 3,000 career MLB hits.

Team	Hits
Mariners	156
Yankees	311
Mariners	2,533

HITS

SOURCE Baseball-reference.com

AP Photo

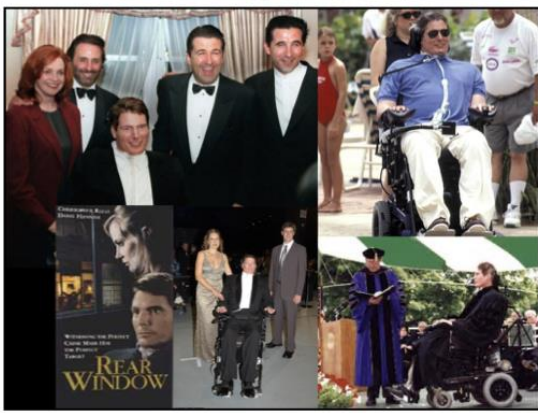
INTERNATIONAL HIT

4,000
CAREER HITS

4,000 HITS
Congrats, Ichiro!



Text 4: Christopher Reeve



**APPENDIX E: NARRATION SCRIPT FOR THE ORAL INTRODUCTION FOR THE
READING TEXT**

Text 1: Steve Jobs 1

1. Steve Jobs was a co-founder and CEO of Apple. He was known as one of the most famous innovators in US history.
2. He was born and grew up in California. The house he grew up in was the original site of Apple Computer.
3. After high school, he entered Reed College in Portland, Oregon. After attending one semester, he dropped out of college because he did not want to spend his parents' money on an education that seemed meaningless to him.
4. However, he started to drop in on calligraphy classes as an auditing student.
5. He learned many typefaces and proportionally spaced fonts in the classes. What he learned in the calligraphy classes had a great impact on his later life. But at the time, he just enjoyed and was fascinated by the beauty of calligraphy.
6. In 1967, Steve Jobs started a company named Apple Computer with his best friend Steve Wozniak at Jobs' parents' garage.
7. In 10 years, Apple had grown from the garage to one of the most valuable companies in the US.

Text 2: Steve Jobs 2

1. In 1983, Steve Jobs hired a new Apple's CEO John Scully, who he thought was very talented to run the company, as a new CEO of Apple.

2. To convince John Scully to come to Apple from Pepsi-Cola, Steve Jobs said, “Do you want to spend the rest of your life selling sugared water, or do you want a chance to change the world?”
3. However, Scully’s and Jobs’ visions of the future of the company began to differ, and eventually, Apple was taken over by Scully. Jobs was technically fired by Apple and then left the company.
4. After leaving Apple, Steve Jobs started a company called NeXT. He also funded Pixar and contributed to the production of *Toy Story*, which turned out to be the most successful animation film in history. Steve Jobs was the executive producer of the film.
5. While Jobs had huge success at Pixar, Apple was struggling and kept losing profits. When Apple was about to go bankrupt, Steve Jobs was asked to return to Apple to rescue the company.
6. After he returned to Apple, Apple released many innovative products: iMac, iPod, iPhone, MacBook Air, and iPad, among others.
7. After introducing numerous innovative products, he was recognized as one of the most innovative businessmen in US history. However, he died of cancer in 2011 at the age of 56.

Text 3: Ichiro Suzuki

1. Ichiro Suzuki is a former Japanese professional baseball player who played both in Japan and the United States. During his professional career, he achieved many impressive titles and received multiple awards in both countries.
2. He began his professional baseball career in Japan in 1992.
3. After playing the first 9 seasons in Japan, he moved to the US and started to play with the Seattle Mariners of Major League Baseball. He achieved multiple titles, including his 10

consecutive 200-hit seasons, the Major League's highest single-season hits with 262 hits, the 3000 hits of his Major League career, and the 4367 total hits in his professional career across Japan and the US.

4. Along with these impressive records in his career, he also impressed many people with the length of his career. Until he retired at the age of 45, he continued to produce high-hitting percentages.
5. He was also known for his strict work ethic, coming to the stadium early to engage in the same stretching and training routines before each game. He had never skipped these routines in his career.
6. To improve his performance while avoiding major muscle-related injuries, he also used special training machines, which may have greatly contributed to his success.

Text 4: Christopher Reeve (Adopted from Izumi et al., 1999)

1. Christopher Reeve was an American actor best known for his role as Superman.
2. He appeared in several successful films after Superman.
3. However, he broke his neck when he was thrown from a horse during a horseback riding competition. The injury paralyzed him from the shoulders down, and he was forced to use a wheelchair for the rest of his life.
4. Despite the accident, he did not give up his hope.
5. With the help of his wife and her encouragement, he returned to his creative work, directing a film and appearing in television series. Over the course of his career, he received multiple awards.
6. He also actively raised money and founded a charitable organization for spinal injury research.

7. Now, he is recognized as a real-life superhero.

APPENDIX F: READING TEXTS

Text 1: Steve Jobs 1

Steve Jobs was known as one of the greatest innovators in history. In his life, he followed an unusual road to success and had several turning points. If he had not dropped out of his college, he would not have dropped in on the calligraphy class. If he had not learned calligraphy, the first Mac computer would not have had wonderful fonts with beautiful calligraphy. When he was 21, he started *Apple* with his best friend Steve Wozniak (*Woz*) in his parent's garage. Steve Jobs would not have started *Apple* if he had not met *Woz*. *Apple* would not have become one of the major computer companies if he had not started *Apple* with *Woz*. (115 words)

Text 2: Steve Jobs 2

When Steve Jobs was 30, he was fired by John Scully, whom he hired as the president of *Apple*. If he had not hired John, Steve Jobs would not have been fired from *Apple*. After he left *Apple*, he did not give up his dreams and started a company named *NeXT* and funded *Pixar*. If he had not gotten fired from *Apple*, *Toy Story* would not have been created. He would not have been able to return to *Apple* again if he had lost his passion. Since then, *Apple* has released many innovative products like iPod, iPhone, iPad, and MacBook Air. These products would not have been created if Steve Jobs had not been fired from *Apple*. (117 words)

Text 3: Ichiro Suzuki

Ichiro Suzuki began his career in Japan and moved to the US in 2001. If he had not moved to the US, he would not have inspired people in both countries. He would not have won many titles in the Major League if he had stayed in Japan. Over the course of his career, he impressed people because he came to the stadium early to follow the same stretching routines before each game. If he had skipped these routines at some point, he might have had a muscle-related injury. He was

also famous for using unique training machines. He would not have become the best hitter in baseball history if he had not used these special training machines. (118 words)

Text 4: Christopher Reeve (Adopted from Izumi et al., 1999)

In 1995, Christopher Reeve fell off his horse. The accident left him paralyzed. If the horse had jumped over the hurdle successfully, Reeve would not have fallen. If his hands had been free, he would have landed safely. Despite the accident, he did not give up his hope to return to creative work and founded a charitable organization for spinal injury research. He would have given up all hope to live if his wife had not encouraged him to be strong. If he had felt discouraged, he would not have recognized his ability to raise money for medical research. Now, he is remembered by his fans as a real-life superman. (110 words)

APPENDIX G: ALL THE EXEMPLAR SENTENCES AND AOIS

Text 1: Steve Jobs 1

Exemplar Sentence (ES) #1 (AOI #1 & 2)

If he had not dropped out of his college, he would not have dropped in on the calligraphy

AOI #1

AOI #2

class.

ES #2 (AOI #3 & 4)

If he had not learned calligraphy, the first Mac computer would not have had wonderful fonts

AOI #3

AOI #4

with beautiful calligraphy.

ES #3 (AOI #5 & 6)

Steve Jobs would not have started *Apple* if he had not met Woz.

AOI #5

AOI #6

ES 4 (AOI #7 & 8)

Apple would not have become one of the major computer companies if he had not started

AOI #7

AOI #8

Apple with Woz.

Text 2: Steve Jobs 2

ES 5: If he had not hired John, Steve Jobs would not have been fired from *Apple*.

ES 6: If he had not gotten fired from *Apple*, *Toy Story* would not have been created.

ES 7: He would not have been able to return to *Apple* again if he had lost his passion.

ES 8: These products would not have been created if Steve Jobs had not been fired from *Apple*.

Text 3: Ichiro Suzuki

ES 9: If he had not moved to the US, he would not have inspired people in both countries.

ES 10: He would not have won many titles in the Major League if he had stayed in Japan.

ES 11: If he had skipped these routines at some point, he might have had a muscle-related injury.

ES 12: He would not have become the best hitter in baseball history if he had not used these special training machines.

Text 4: Christopher Reeve (Adopted from Izumi et al., 1999)

ES 13 (AOI #1 & 2)

If the horse **had jumped** over the hurdle successfully, Reeve **would not have fallen**.

AOI #1

AOI #2

ES 14 (AOI #3 & 4)

If his hands **had been** free, he **would have landed** safely.

AOI #3

AOI #4

ES 15 (AOI #5 & 6)

He **would have given** up all hope to live if his wife **had not encouraged** him to be strong.

AOI #5

AOI #6

ES 16 (AOI #7 & 8)

If he **had felt** discouraged, he **would not have recognized** his ability to raise money for medical

AOI #7

AOI #8

research.

APPENDIX H: WRITTEN PICTURE-CUED DESCRIPTION TEST

Practice (Example) Task

Please read the following passage.



Michael Jackson

Michael Jackson, an American singer, was often called the “King of Pop.” He died in 2009 at the age of 50, but he is still popular around the world. When he was five, he started to sing professionally. So, When he was a child, he never had a good childhood. Thus, he came to believe that every child should have a good childhood.

Example Picture 1

1960'

When

<p>Michael Jackson</p>  <p>5 years old</p>	<p>professional</p> 
BE	START TO SING

- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **“When ...”**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.



Please write your one-sentence description here.

A model description: **When Michael Jackson was five years old, he started to sing professionally.**

Practice Picture 2

1960'

When

<p>Michael Jackson</p> 	<p>A good childhood</p> 
<p>BE</p>	<p>HAVE</p>

- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"When ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.
- (5) From this time, you can use "he" instead of spelling out "Michael Jackson".

Please write your one-sentence description here.

A model description: When he was a child, he never had a good childhood.

This is the end of the practice (example) section.

Now, you are going to work on the real written description tasks from Part 1 to Part 3.

Let's start from Part 1 as you worked on the practice tasks.



Part 1

Please read the following passage.

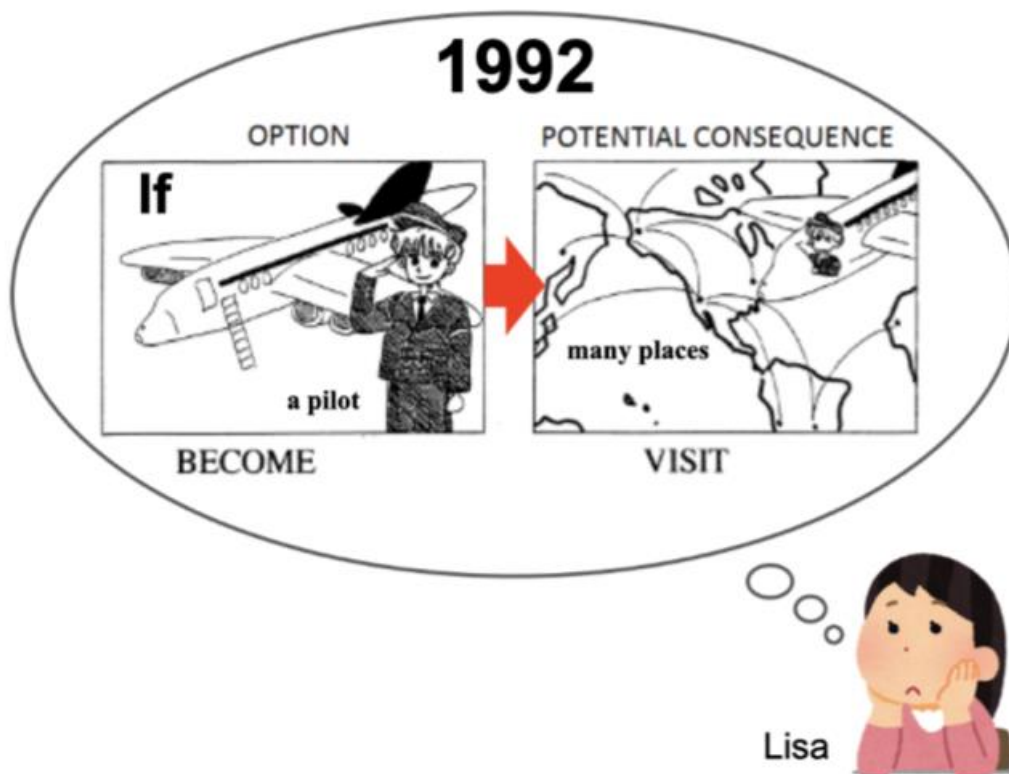
Lisa

Lisa had many options when she finished high school in 1992. However, she decided to work for an insurance company. The following pictures illustrate Lisa's other options and their potential consequences (results). She did not choose these options.

Instruction

In each pair, Picture A shows Lisa's option when she finished high school in 1992, and Picture B shows the potential consequence (result) of the option.

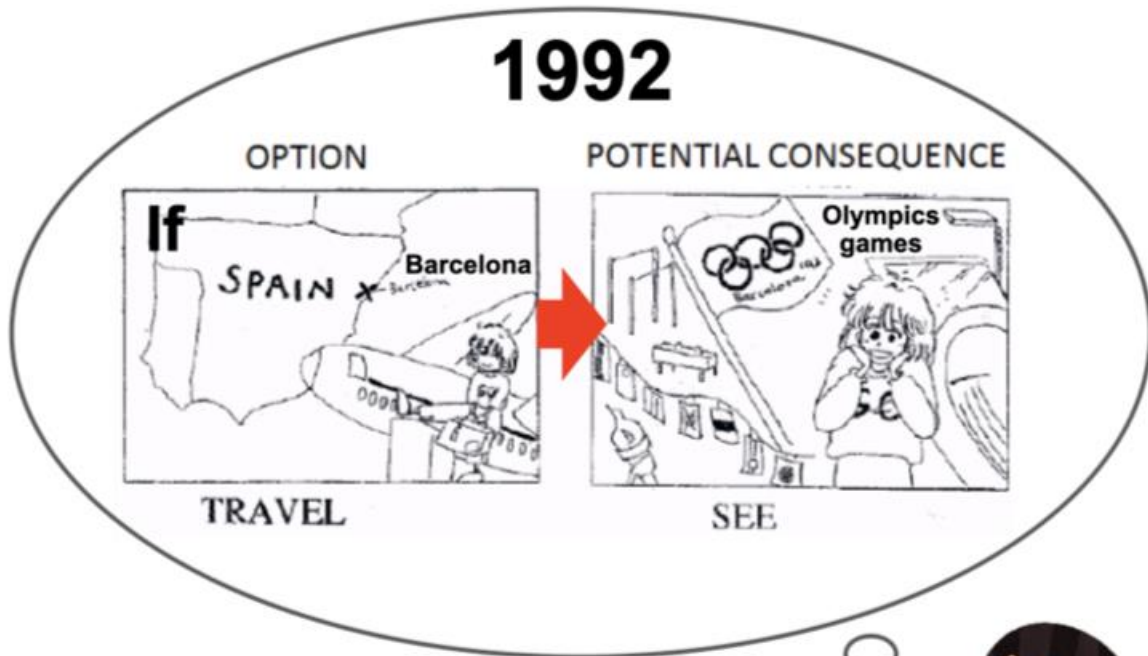
Picture 1



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.

Please write your one-sentence description here.

Picture 2





- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.

Please write your one-sentence description here.

Picture 3

1992

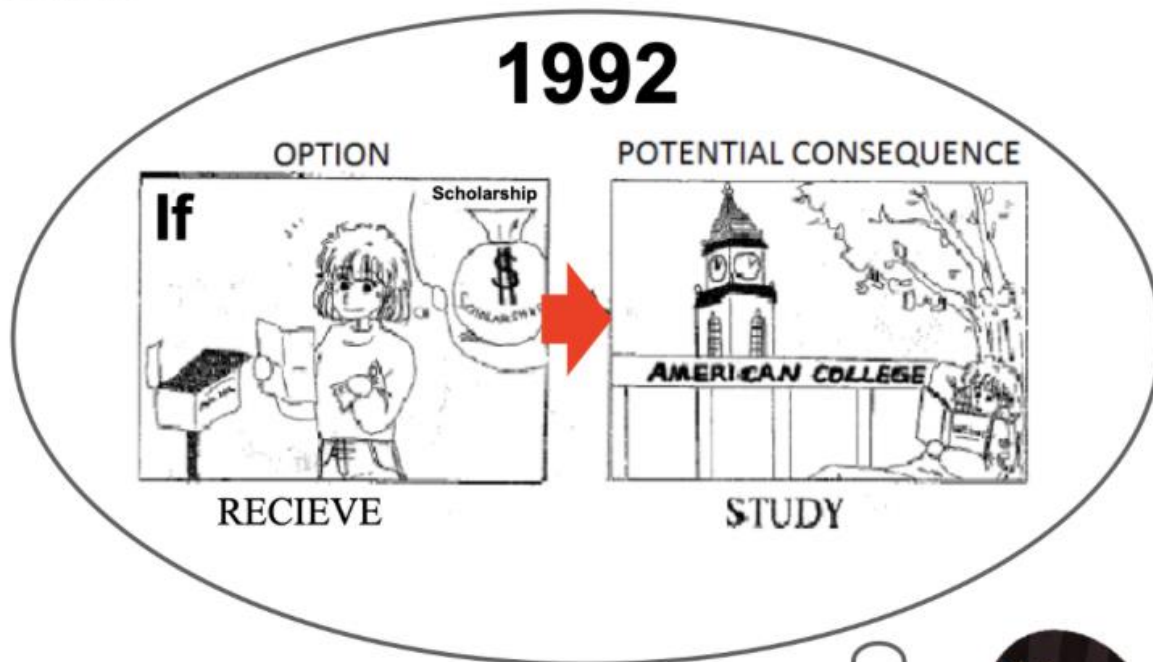
OPTION	POTENTIAL CONSEQUENCE
<p>If</p>  <p>Josef</p>	
GET MARRIED	MOVE



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.

Please write your one-sentence description here.

Picture 4



Lisa

- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form, if necessary.

Please write your one-sentence description here.

This is the end of Part 1. Let's move on to Part 2.



Part 2 (5 pictures)

Please read the following passage.

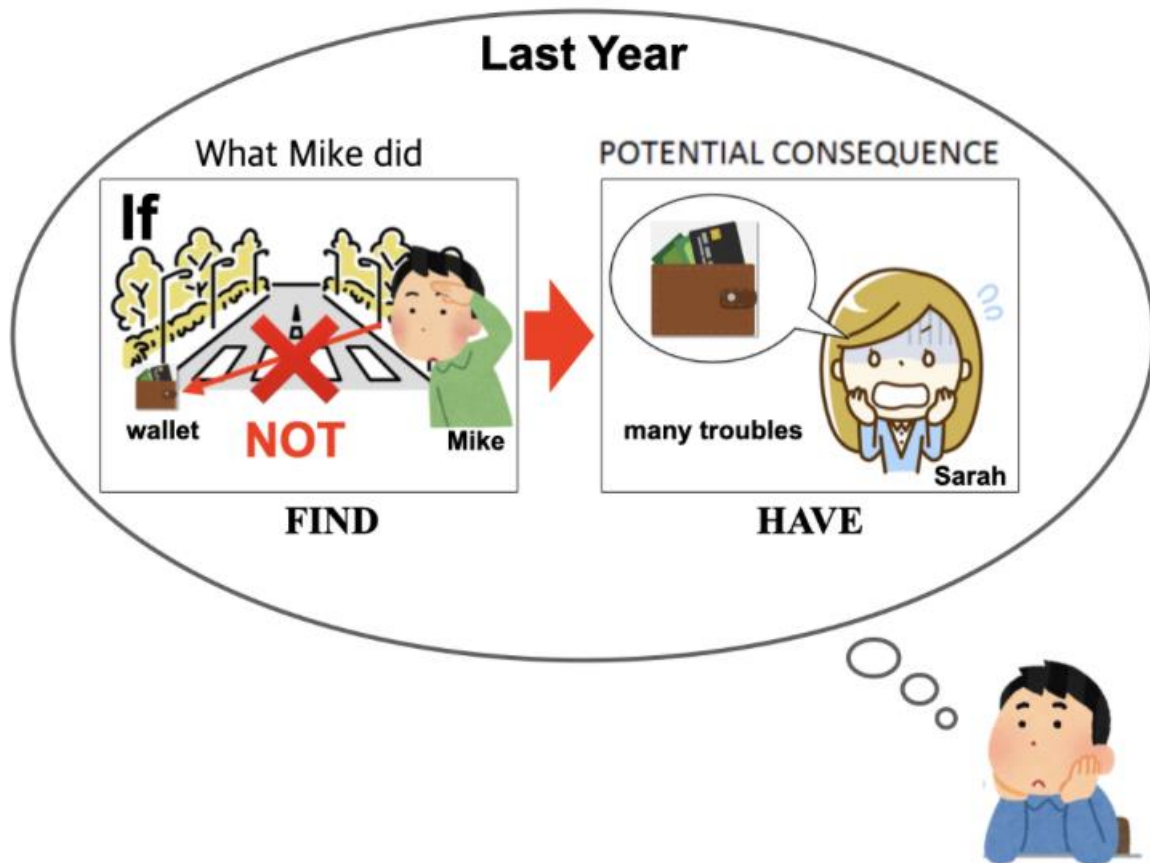
Mike and Sarah

Mike found a missing wallet the other day and gave it to the police at the police station. At the same time, Sarah was reporting her missing wallet at the same police station. She was very thankful that Mike had returned her wallet because she had several valuable items in the wallet. A few days later, Sarah asked Mike out for dinner at a nice restaurant to express her appreciation. Both of them enjoyed the dinner and started dating. The next year, they got married and lived happily ever after.

Instruction

In each pair, Picture A shows what Mike or Sarah did or their options in the story, and Picture B shows the potential consequence (result) for them. All the events happened last year.

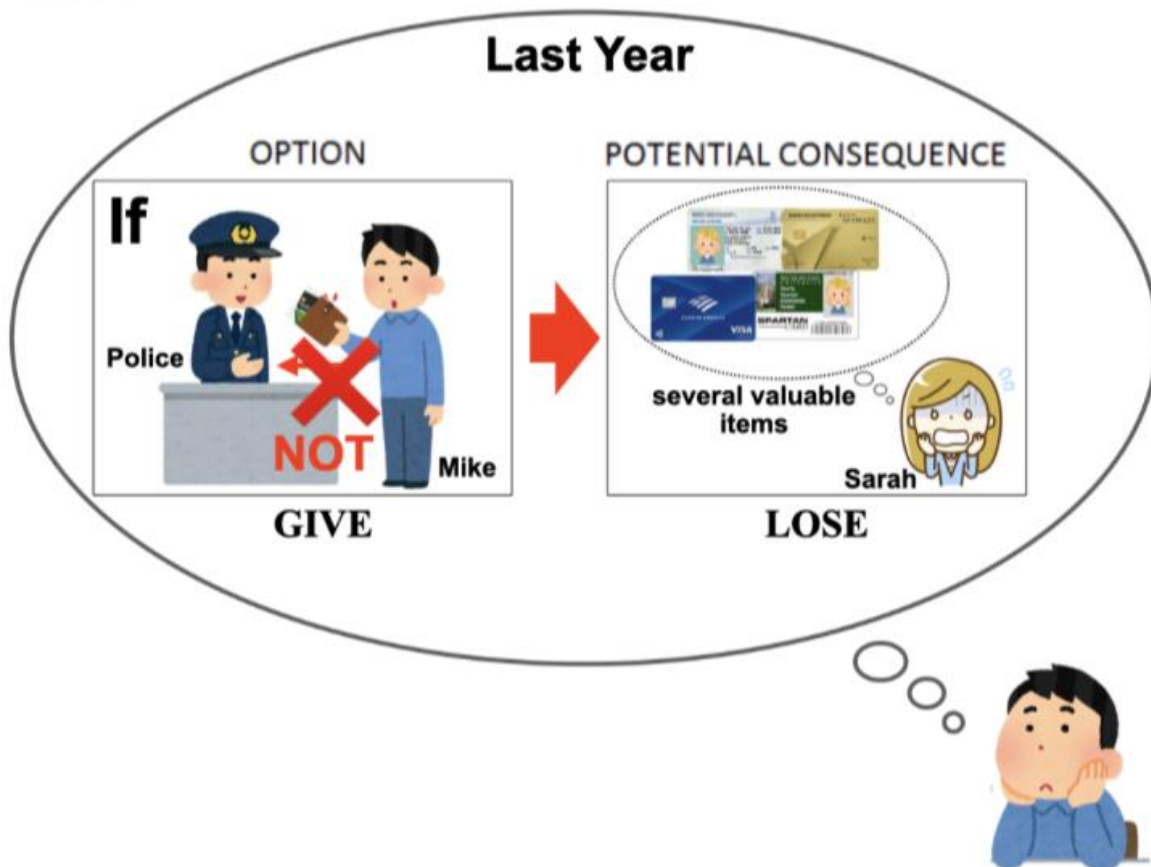
Picture 5



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **“If ...”**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

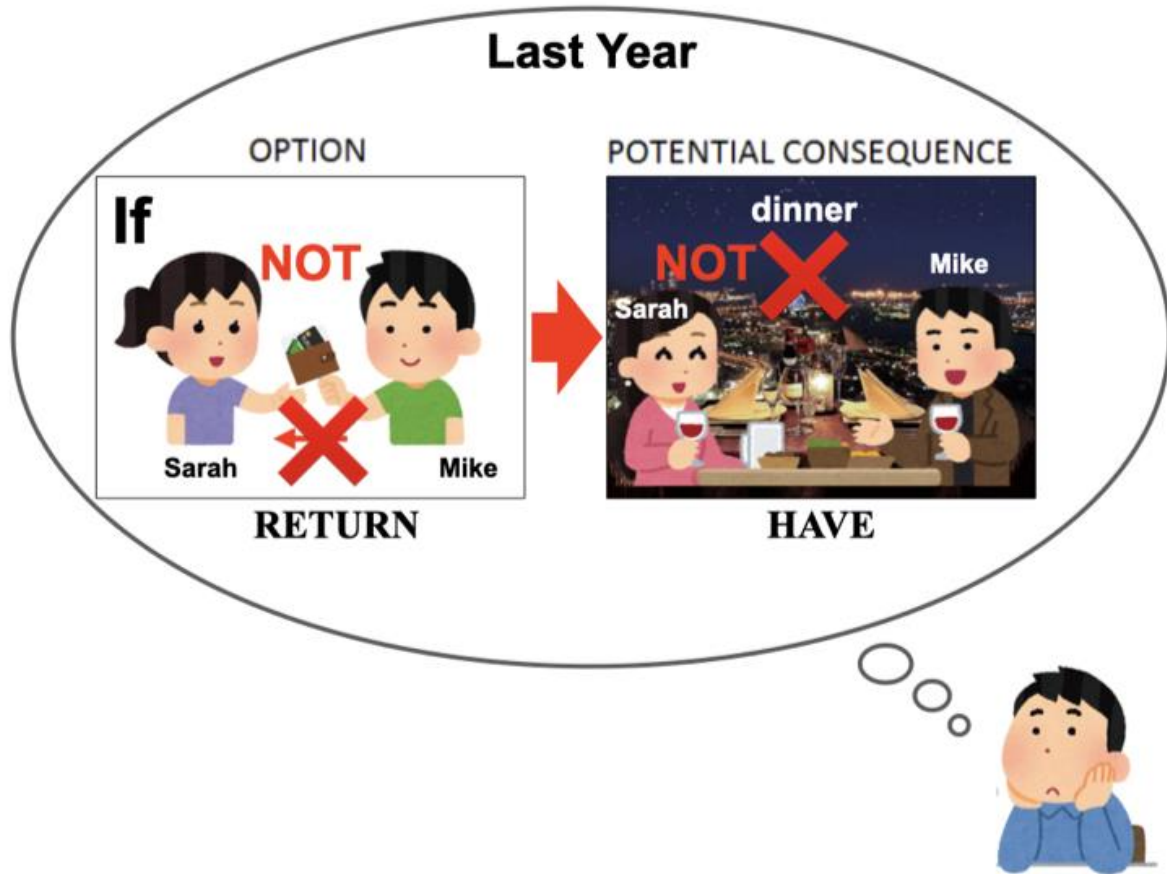
Picture 6



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **“If ...”**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

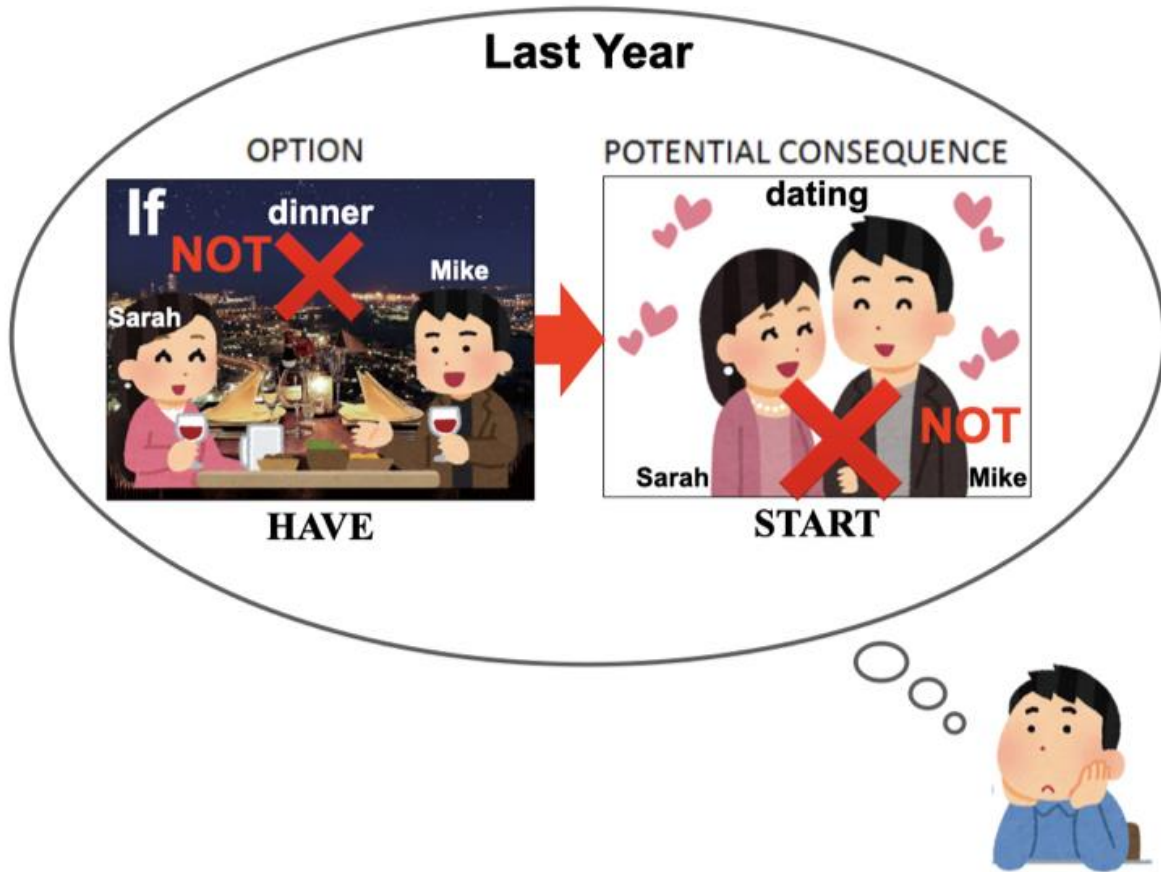
Picture 7



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **“If ...”**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

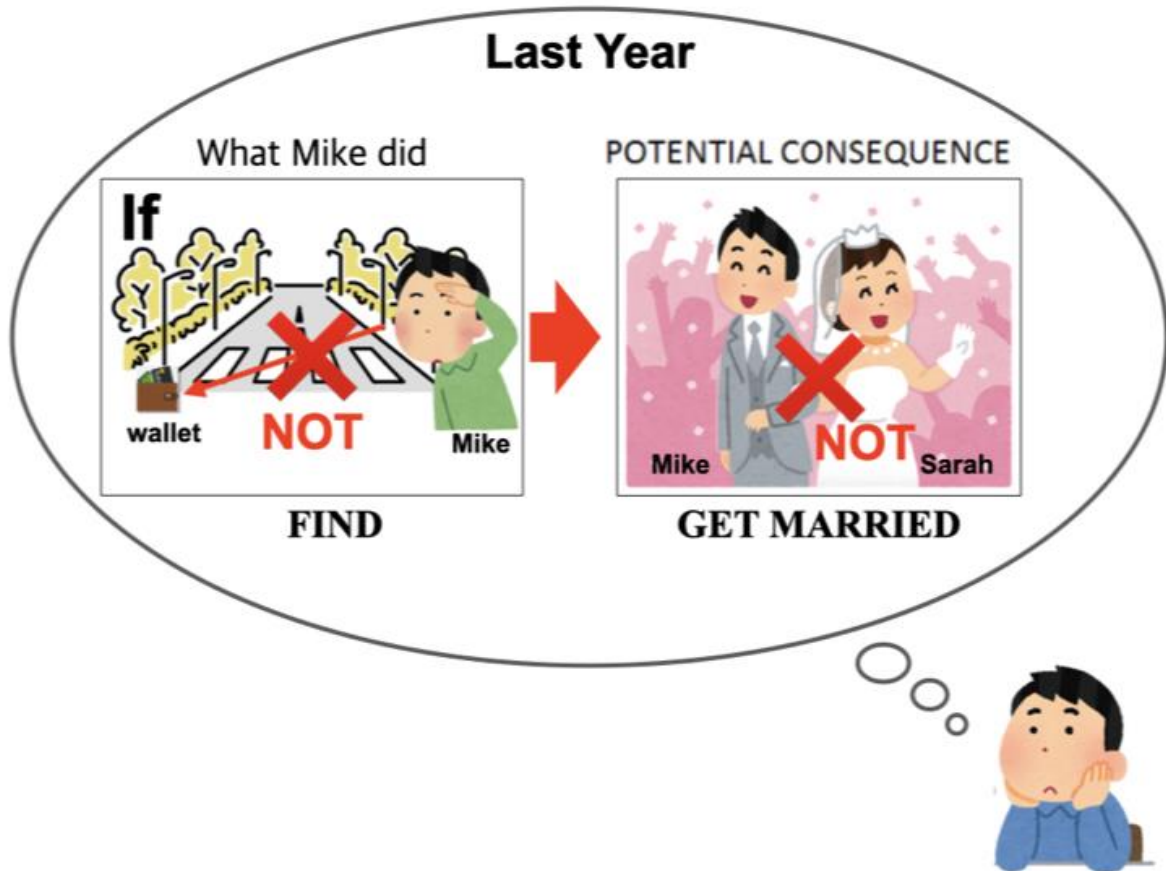
Picture 8



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

Picture 9



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

This is the end of Part 2. Let's move on to the final part, Part 3.



Part 3 (5 pictures)

Please read the following passage.

Michael J. Fox

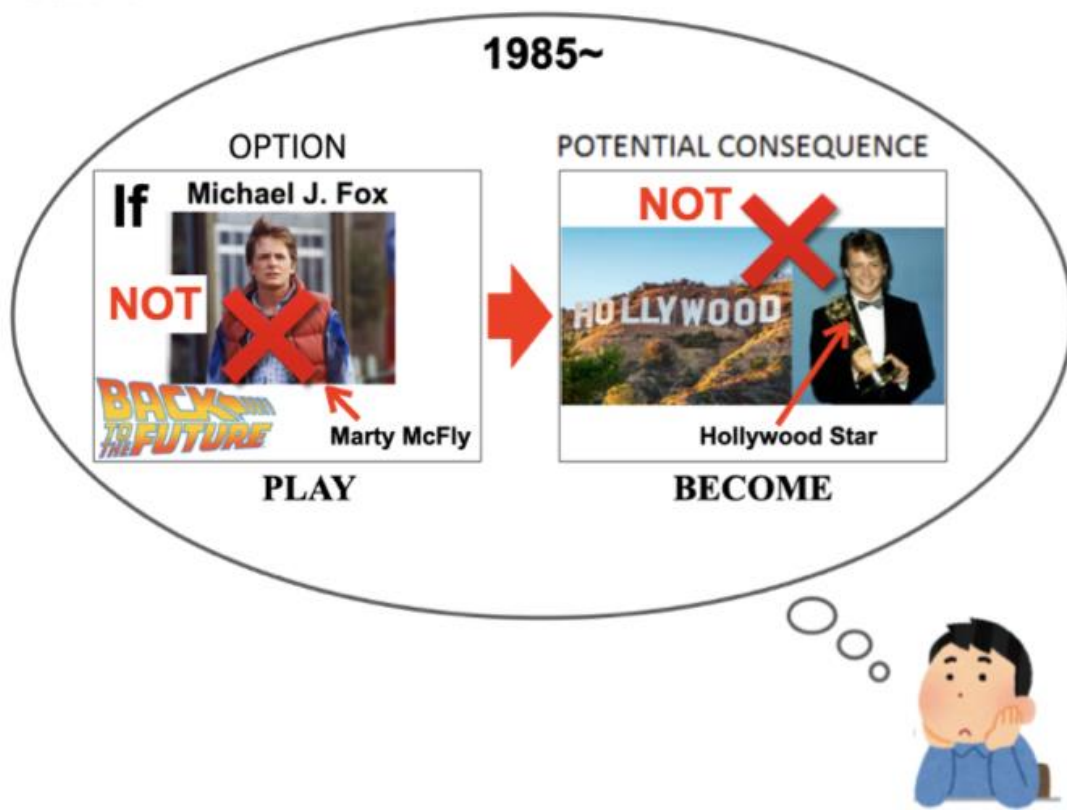
A Canadian-American actor Michael J. Fox became famous for his role as Marty McFly in Back to the Future (1985-1990). He became one of the most popular Hollywood stars after the movie. Michael J. Fox had Parkinson's disease when he was 29 years old. After that, he reduced his activities as an actor. Despite this challenge, he did not give up hope. He always saw his challenges as opportunities. He continued to make guest appearances on TV and worked as a voice actor. Furthermore, he wrote several books about his life and encouraged many people who are facing life challenges.

Along with these activities, he also supported medical research to find a cure for Parkinson's disease. In 2000, he started the Michael J. Fox Foundation to raise money for research.

Instruction

In each pair, Picture A shows what happened to Michael J. Fox or his options, and Picture B shows the potential consequence (result) for them.

Picture 10




- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.

Please write your one-sentence description here.

Picture 11

1991~

What happened to him

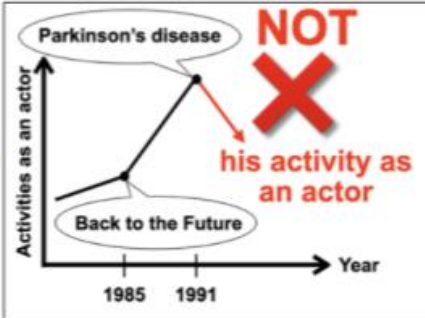


IF **NOT**

Michael J. Fox

HAVE


POTENTIAL CONSEQUENCE



NOT

his activity as an actor

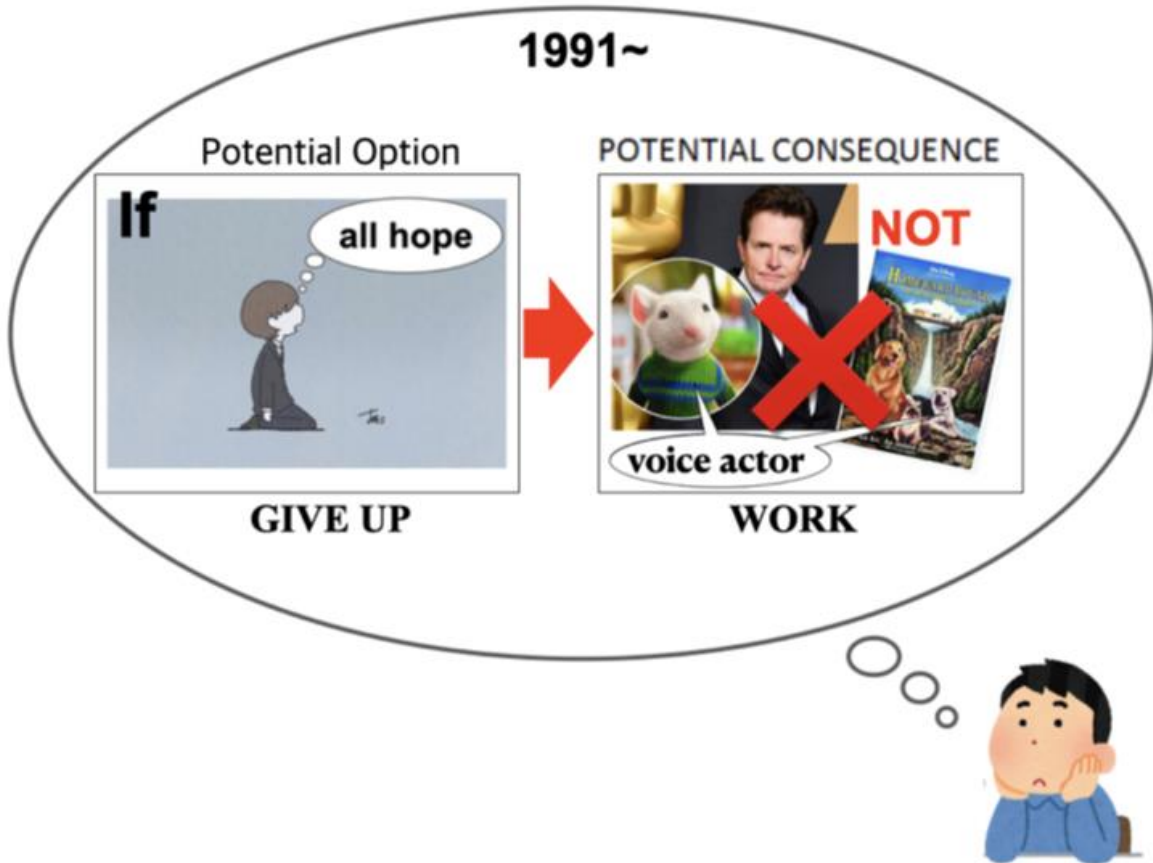
REDUCE



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.
- (5) From this time, you can use "he" instead of spelling out "Michael J. Fox".

Please write your one-sentence description here.

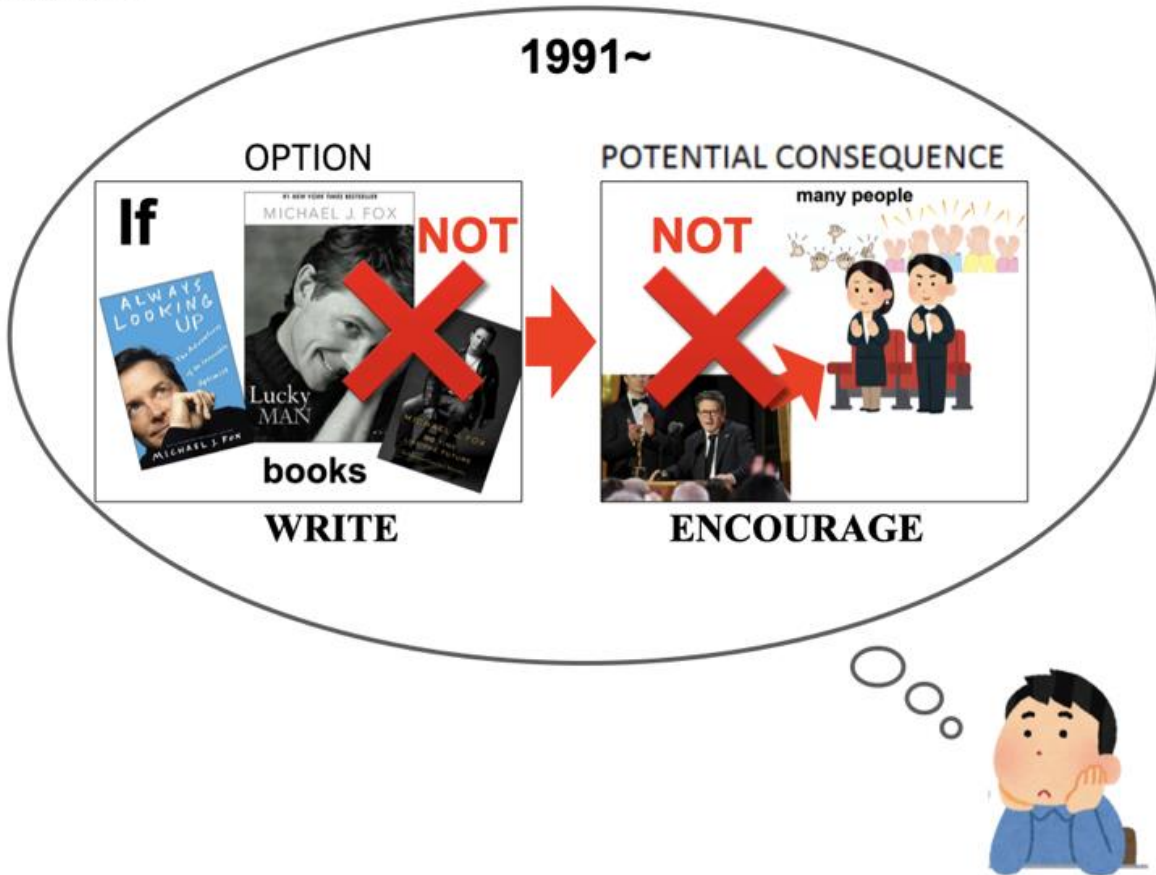
Picture 12



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.
- (5) You can use "he" instead of spelling out "Michael J Fox".

Please write your one-sentence description here.

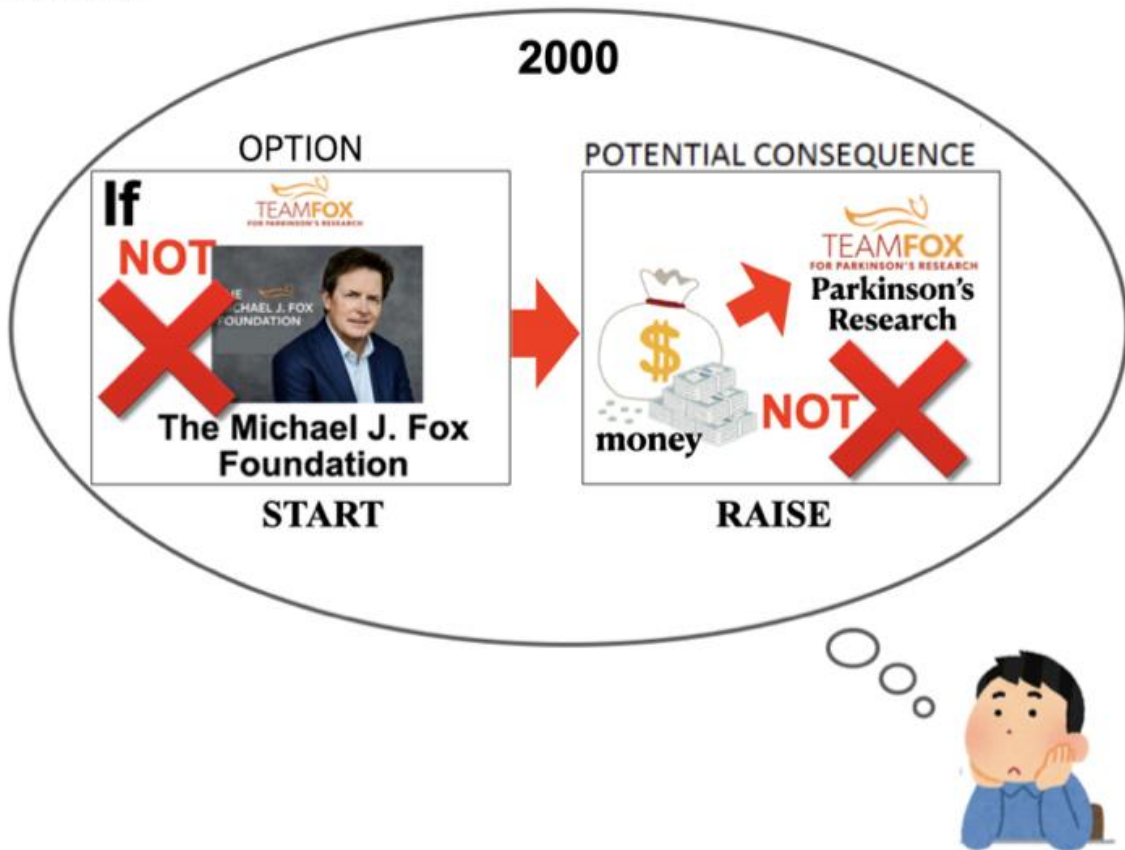
Picture 13



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.
- (5) You can use "he" instead of spelling out "Michael J Fox".

Please write your one-sentence description here.

Picture 14



- (1) Describe the two pictures in **one sentence**.
- (2) Start each sentence with **"If ..."**
- (3) **Use the verb** given below each picture when you make a sentence.
- (4) Change the verb form if necessary.
- (5) You can use "he" instead of spelling out "Michael J Fox".

Please write your one-sentence description here.

This is the end of the written picture description task.

Thank you very much for your participation.



APPENDIX I: ALL THE OEIT AND THE DETAILS FOR EACH TEST ITEM

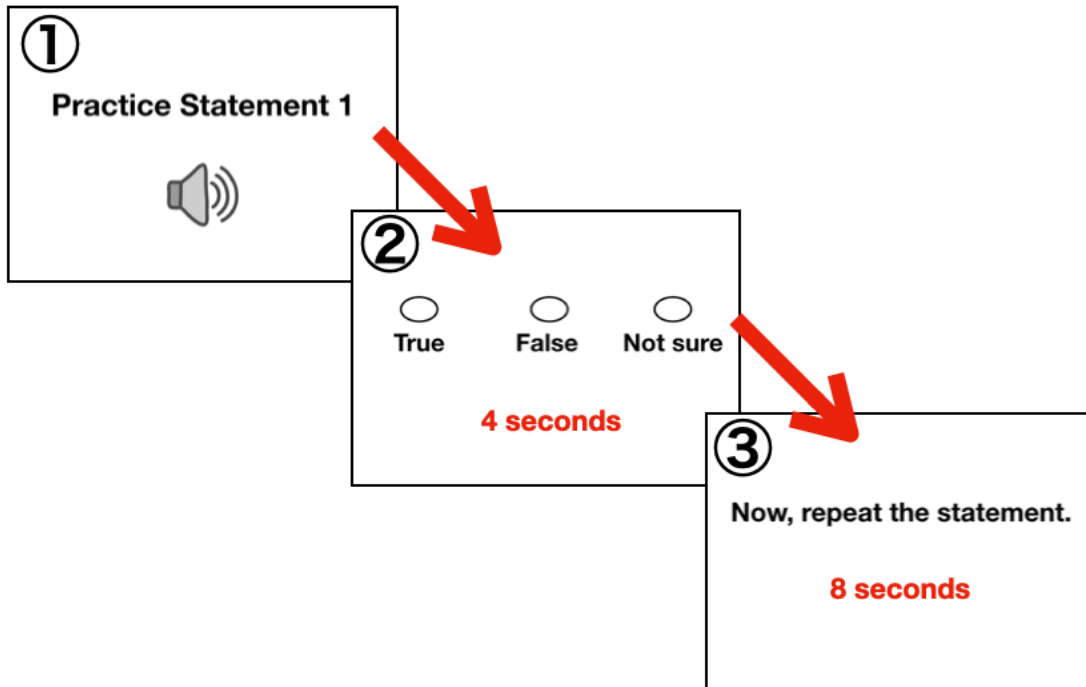
Table 7.1. *All the OEIT and the details for each test item*

<i>Item #</i>	<i>New Item ID</i>	<i>Old Item ID</i>	<i>Statements</i>	<i># of syllables</i>	<i>Syllable length</i>	<i>Sentence length</i>	<i>Audio length</i>
1	PHC1	PHC2	If I had had more time during my high school years, I would have spent more time on my hobby.	22	Long	20	8.32
2	PHC2	PHC3	If I had not studied English hard, I would have given up studying in the US.	21	Long	16	7.21
3	PHC3	PHC4	If the government had not provided free COVID vaccinations, more people would have died.	23	Long	14	8.32
4	PHC4	PHC5	If Michael Jackson had not become famous in his childhood, he might have lived a happy life.	23	Long	17	8.2
5	PHC5	PHC6	If I had known today's inflation, I would not have come to the US.	17	Long	14	6.48
6	PHC6	PHC8	If Obama had married a different woman, he would not have become president.	22	Long	13	7.77
7	PHC7	PHC9	If Edison had not invented electricity, he could not have changed the world.	21	Long	13	7.12
8	PHC8	PHC11	I might have had better grades in high school if I had studied math harder.	19	Long	15	6
9	PHC9	PHC12	I could have received a college scholarship if I had prepared earlier.	19	Long	12	5.92
10	PHC10	PHC13	I would have entered a different university if I had known the winter in Michigan.	25	Long	15	6.79
11	PHC11	PHC16	My childhood might not have been happy if I had not lived with a dog.	17	Long	15	6.19
12	PHC12	PHC17	I could not have been able to speak English if I had not met a nice English teacher.	22	Long	18	7.48
13	PHC13	PHC18	Obama would not have become president if he had not studied at Harvard.	20	Long	13	6.55
14	PHC14	PHC20	Mariah Carey could not have become a celebrity if she had not written the Christmas song.	24	Long	16	7.84
15	D1	D3	Human beings are the only animal that can use language.	16	Long	10	5.2
16	D2	D4	One thing that you can do to stay healthy is not eat McDonald's.	16	Long	13	5.92
17	D3	D5	Global warming is an issue that all Americans must think about.	19	Long	13	6.36
18	D4	D6	To be successful, you need to be able to speak English.	15	Medium	11	5.37
19	D5	D7	The number of Korean students is increasing in American universities.	21	Long	10	5.97
20	D6	D8	Usually, school teachers make their students study hard.	14	Medium	10	5.68
21	D7	D9	Quitting smoking isn't easy for people who are addicted.	13	Medium	9	5.54

Table 7.1 (cont'd)

<i>Item #</i>	<i>New Item ID</i>	<i>Old Item ID</i>	<i>Positive/Negative</i>	<i>Subject in MC</i>	<i>TL form</i>	<i>Grammaticality</i>	<i>Other than K1-K2 words</i>	<i>Sentence Type</i>
1	PHC1	PHC2	Positive	1st	Past Hypothetical Conditional	G		If Clause First
2	PHC2	PHC3	Positive	1st	Past Hypothetical Conditional	G		If Clause First
3	PHC3	PHC4	Positive	3rd	Past Hypothetical Conditional	G	vaccinations-6	If Clause First
4	PHC4	PHC5	Positive	3rd	Past Hypothetical Conditional	G		If Clause First
5	PHC5	PHC6	Negative	1st	Past Hypothetical Conditional	G	inflation-3	If Clause First
6	PHC6	PHC8	Negative	3rd	Past Hypothetical Conditional	G		If Clause First
7	PHC7	PHC9	Negative	3rd	Past Hypothetical Conditional	G	invented-3	If Clause First
8	PHC8	PHC11	Positive	1st	Past Hypothetical Conditional	G		Main Clause First
9	PHC9	PHC12	Positive	1st	Past Hypothetical Conditional	G	scholarship-4	Main Clause First
10	PHC10	PHC13	Positive	1st	Past Hypothetical Conditional	G		Main Clause First
11	PHC11	PHC16	Negative	1st	Past Hypothetical Conditional	G		Main Clause First
12	PHC12	PHC17	Negative	1st	Past Hypothetical Conditional	G		Main Clause First
13	PHC13	PHC18	Negative	3rd	Past Hypothetical Conditional	G		Main Clause First
14	PHC14	PHC20	Negative	3rd	Past Hypothetical Conditional	G	celebrity-4	Main Clause First
15	D1	D3	N/A		RC (SU)	G		N/A
16	D2	D4	N/A		RC (DO)	G		N/A
17	D3	D5	N/A		RC (OPREP)	G		N/A
18	D4	D6	N/A		Infinitive	G		N/A
19	D5	D7	N/A		Progressive	G		N/A
20	D6	D8	N/A		Causative	G		N/A
21	D7	D9	N/A		Gerund + RC (SU)	G		N/A

APPENDIX J: AN EXAMPLE DISPLAY FO THE OEIT TEST FORMAT



APPENDIX K: ASSUMPTION CHECKING FOR MIXED DESIGNED ANOVA FOR THE GRAMMAR DEVELOPMENTAL TESTS (OEIT & WPDT)

Assumption checking for conducting the mixed-design ANOVA

The following assumptions for performing mixed-design ANOVAs were checked on the results of the OEIT and the WPDT: the normality of the distribution, homogeneity of the variance, normal distribution of the residuals, and equal variance of the residuals.

OEIT Results

The normal distribution of the data set was examined through the histograms, the qq-plots, and the results of Shapiro-Wilk tests. Based on the histograms, the participants' test-performances were not always normally distributed throughout the entire testing-sessions from pretest to the delayed posttest. However, the qq-plots indicated that the distributions were relatively aligned with the regression lines. As indicated in the visual inspection of the qq-plots, the results of the Shapiro-Wilk tests indicated non-significant values for the data sets except the results of the Written Output group's delayed posttest on past passive ($W = 0.90, p = .01$). Based on these observations, the assumption of the normal distribution of the data was acceptable. As for the homogeneity of variance, the size of the boxes between groups were slightly different but the results of the Levene's test on the pretest results indicated non-significant value ($F(2, 80) = 1.94, p = .15$). Thus, the variance for the pretest were relatively equal between the group at the point of the pretest and the variances were also relatively similar between the groups on the results of the immediate posttest and the delayed posttest.

The distribution of the residuals was extracted from the model of the mixed-design ANOVA. The histogram of the residual appeared to be normally distributed but the results of the Shapiro-Wilk test on the residuals indicated a significant value ($W = 0.96, p < .001$). On the

Residuals vs Fitted plots for the residuals, the residuals were not evenly distributed in the plot. Thus, the homogeneity of residuals was not met as shown in the Residuals vs Fitted plots for the residuals. Therefore, the ANOVA results were carefully interested using both descriptive statistics, 95% CIs, and also together with visual inspections of the box plots in Figure 4.5.

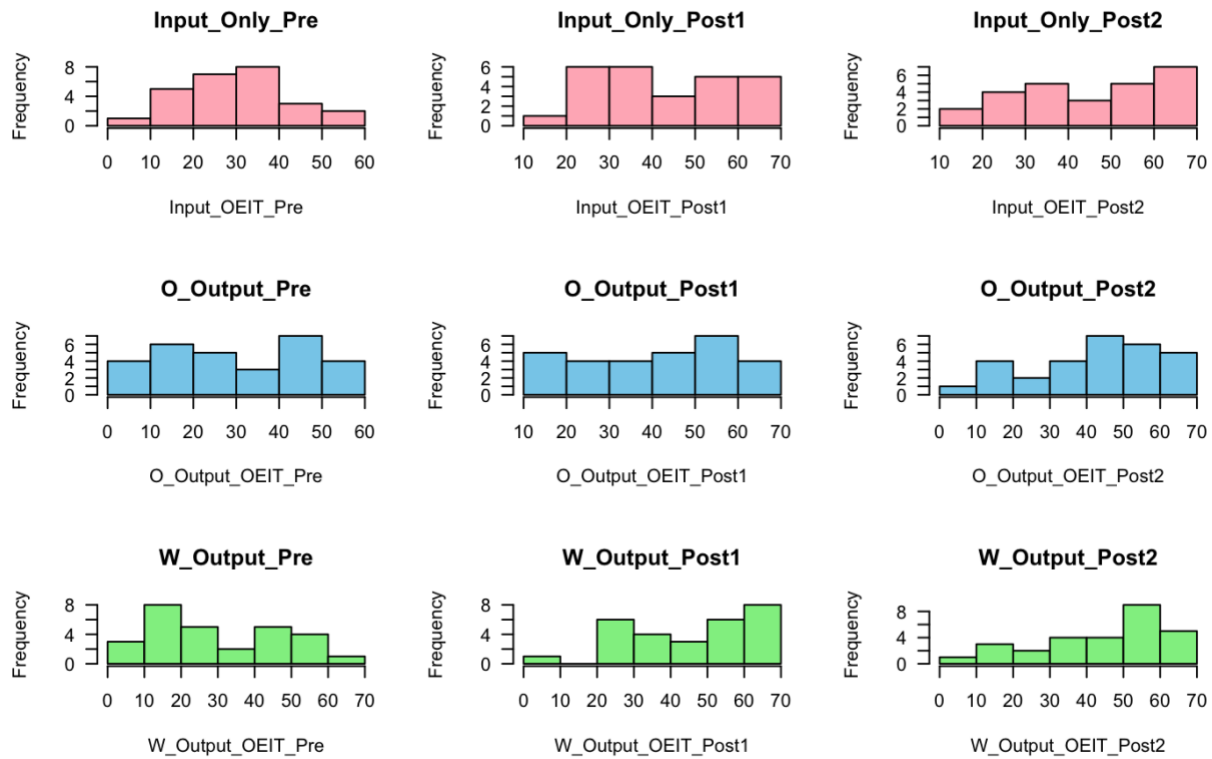


Figure 7.1. Histograms for the distribution of the results of the OEIT

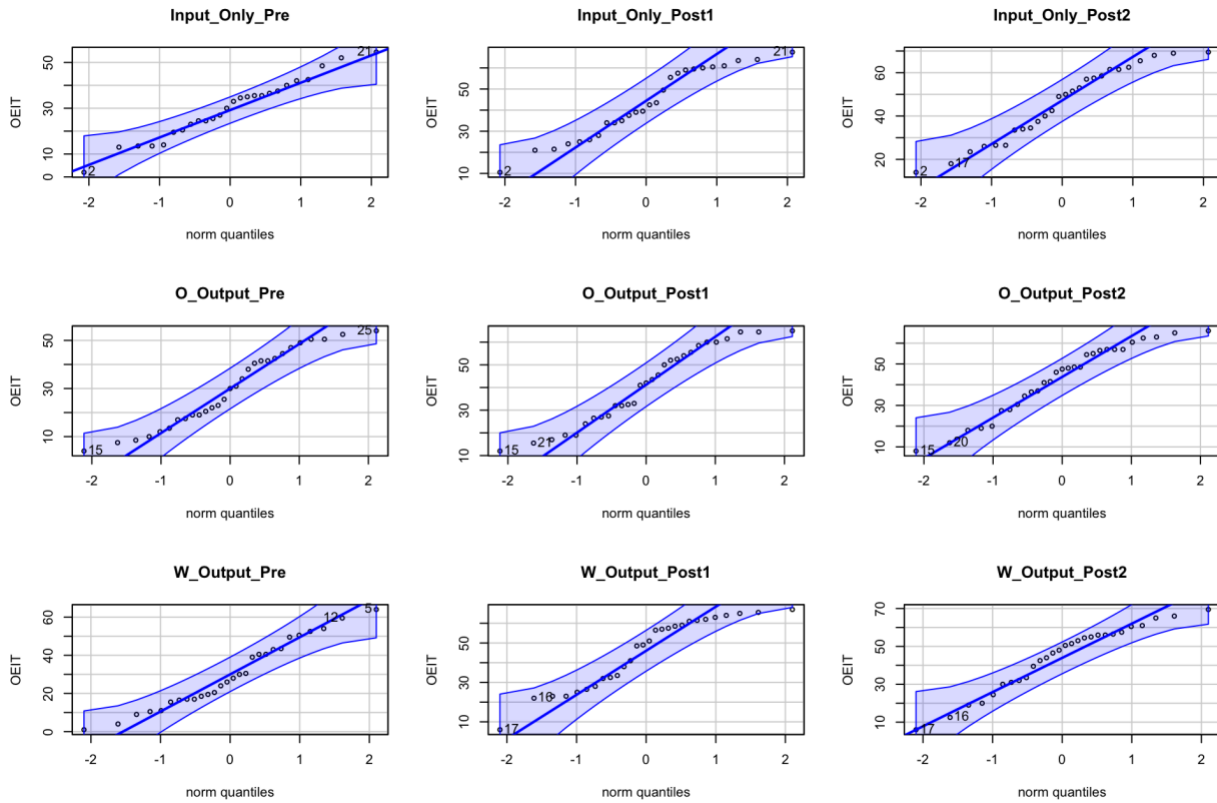


Figure 7.2. The qq-plots for all the test results on the OEIT

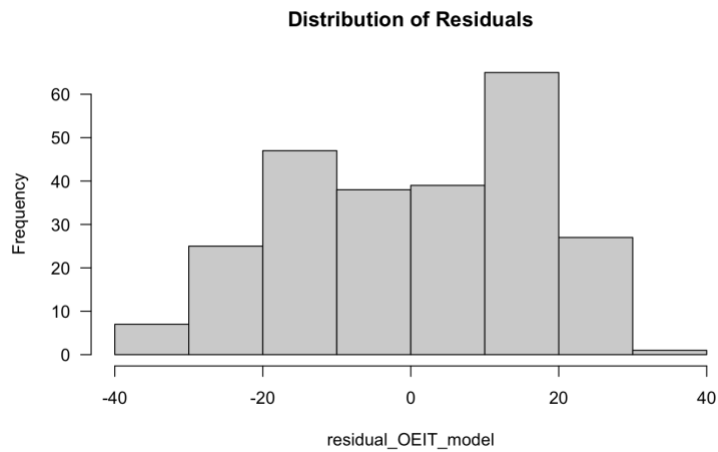


Figure 7.3. Histogram for the distribution of the residuals

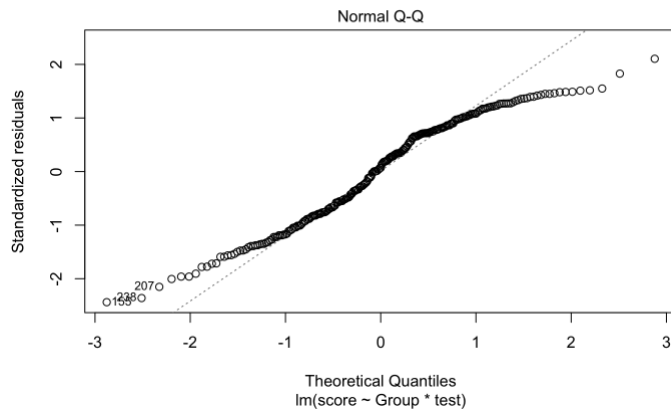


Figure 7.4. A normal qqplot for the residuals (OEIT)

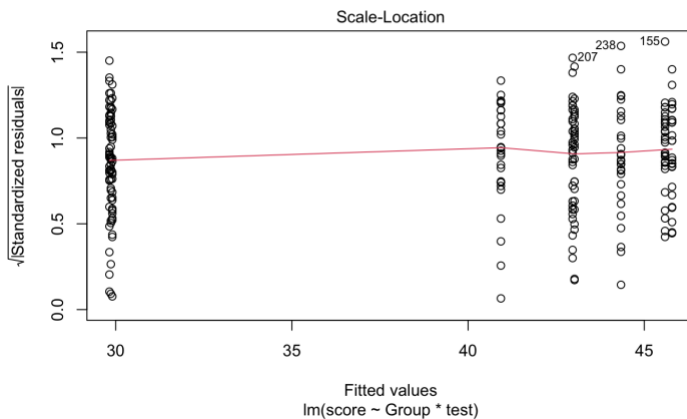


Figure 7.5. Residuals vs Fitted plots for the residuals (OEIT)

WPDT Results

The assumption checking for the results of the WPDT was also conducted as it was done for the results of the OEIT. All the groups' test-performances were not normally distributed and the Written Output group's distributions were negatively distributed. The qq-plots also indicated similar observations, which indicated that the distributions were not aligned along the regression lines particularly as they work on the posttests. This visual observation was also confirmed by the results of the Shapiro-Wilk tests, indicating significant values for most of the data sets. Based on these observations, the assumption of the normal distribution of the data were not met on the results of the WPDT. As for the homogeneity of variance, the size of the boxes of the pretest

were relatively equal between the groups. The results of the Levene's test also indicated non-significant value ($F(2, 80) = .17, p = .82$). The variances also seemed to differ between the groups as the participants worked on the immediate posttest and the delayed posttest (see Figure 4.7).

The distribution of the residuals was extracted from the model of the mixed-design ANOVA. The histogram of the residual appeared to be normally distributed but again, the results of the Shapiro-Wilk test on the residuals indicated a significant value ($W = 0.96, p < .001$). On the Residuals vs Fitted plots for the residuals, the residuals were not evenly distributed in the plot. Thus, the homogeneity of residuals was also not met as shown in the Residuals vs Fitted plots for the residuals. Again, the ANOVA results were carefully inspected using both descriptive statistics, 95% CIs, and also together with visual inspections of the box plots presented in Figure 4.7.

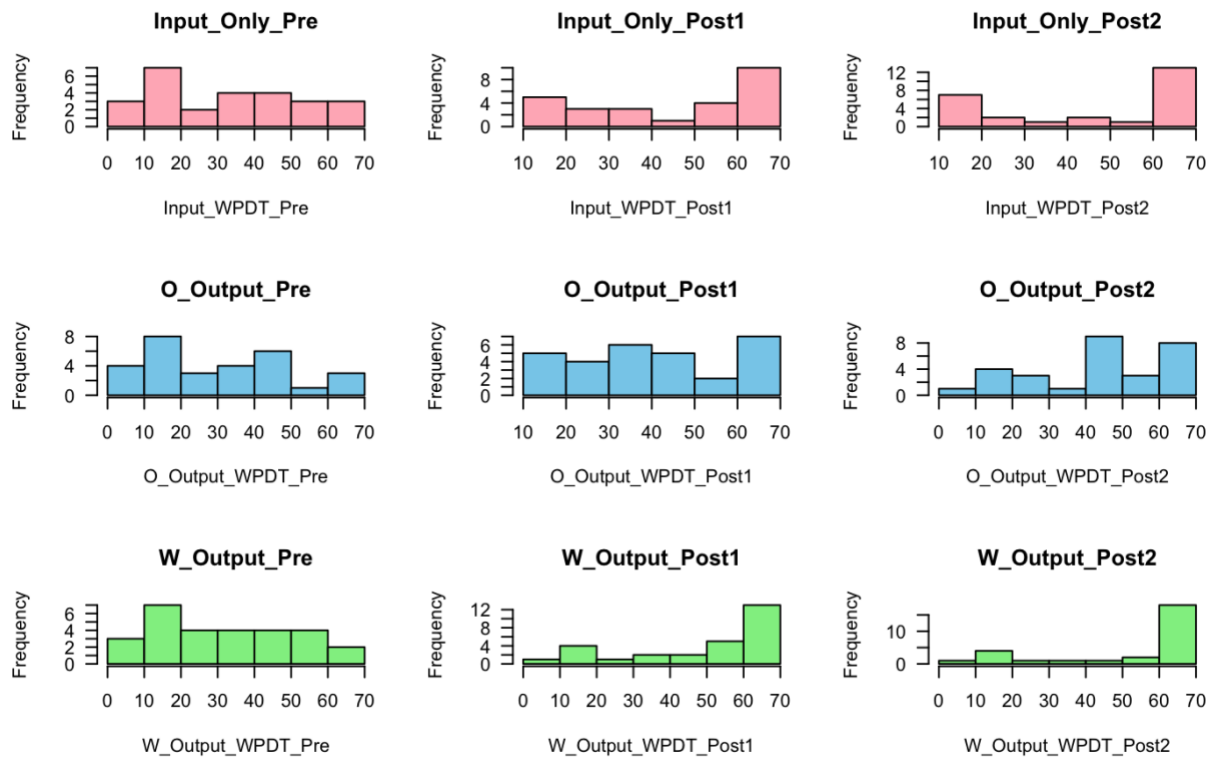


Figure 7.6. Histograms for the distribution of the results on past passive on the WPDT

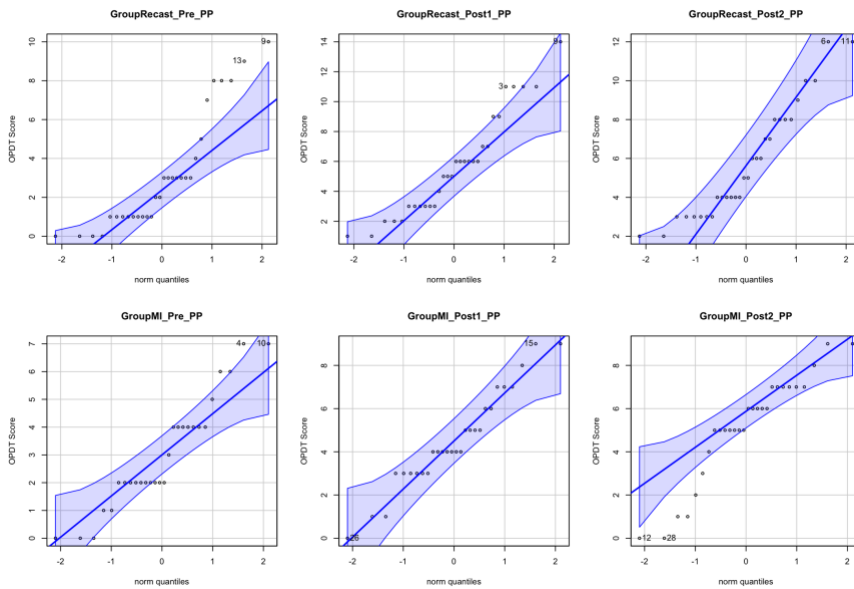


Figure 7.7. The qq-plots for all the test results (WPDT)

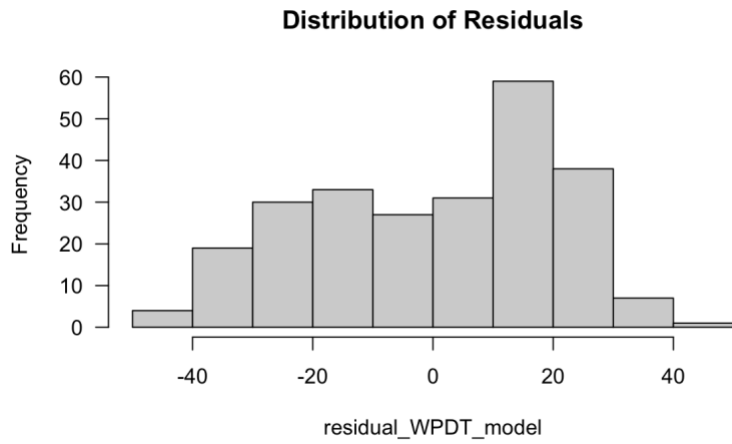


Figure 7.8. Histogram for the distribution of the residuals (WPDT)

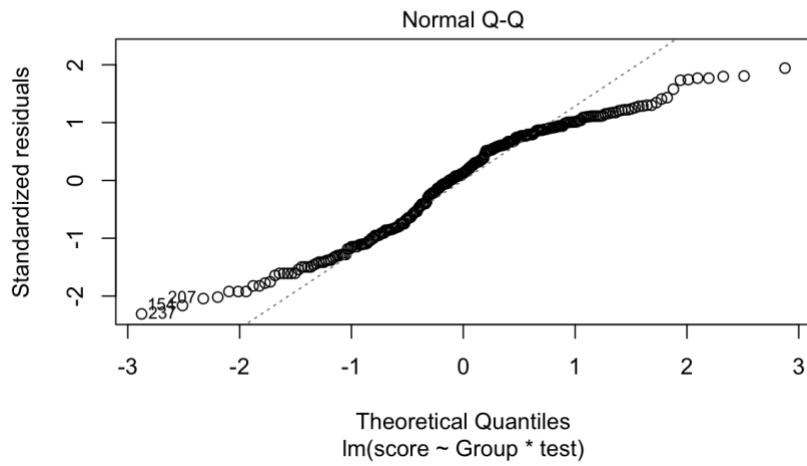


Figure 7.9. A normal qqplot for the residuals (WPDT)

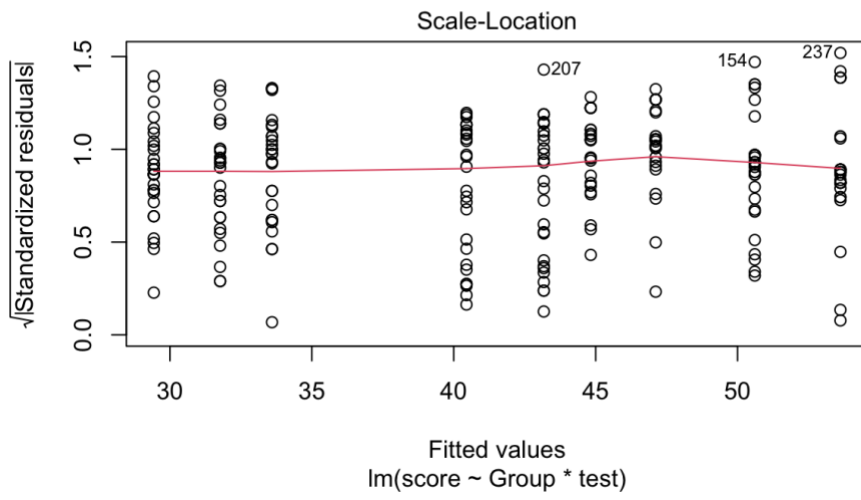


Figure 7.10. Residuals vs Fitted plots for the residuals (WPDT)