ABSTRACT

Title of Document: EFFECTS OF DIFFERENT TYPES OF AUDITORY INPUT ON INCIDENTAL VOCABULARY LEARNING BY L2 JAPANESE SPEAKERS

Kyoko Kobayashi Hillman, Ph.D., 2020

Directed By: Dr. Steven J. Ross, Second Language Acquisition

Research on advanced L2 adult learners and viable classroom instruction for them has become increasingly important along with increasing global connections. This study investigated the effects of different kinds of spoken input modification on listening comprehension and incidental vocabulary learning by 106 advanced Chinese speakers of Japanese in Japan. The participants were randomly assigned to four types of input (genuine, simplified, elaborated, modified elaborated) used in four short academic talks by Japanese professionals. Each talk contained eight low-frequency nouns, each appearing three times. Learning outcomes were assessed using three different measures: form-recognition, meaning recognition with contextual information, and meaning recognition via L2 definitions. Participants responded to three types of comprehension questions (replication, synthesis, inference) while listening to the talks. Scores from an online proficiency test and two working memory (WM) tasks served as moderator variables.

Results showed that elaborated input was the most effective of the four types for both comprehension and incidental vocabulary learning. Results also showed that modified elaborated input, a novel input modification type that contained the same elaboration but with shorter sentences, was more effective when higher WM was available. In contrast, elaborated input was least influenced by WM capacities. Regarding

the relationships between input modification and type of comprehension questions, modified elaborated input had a marginally significant effect on replication items. For synthesis and inference items, statistically significant effects for input type were not found, contradicting previous results in the literature. Proficiency showed significant effects on all tests, whereas WM showed interaction effects with simplified and modified elaborated input.

In light of these findings, the study concludes that (a) elaborated input is more beneficial for advanced L2 learners than genuine input regardless of WM, (b) modified elaborated input with short sentences requires WM, (c) input elaboration is more effective than input enhancement for incidental vocabulary learning for both form and meaning recognition, and (d) enhanced incidental vocabulary conditions using greater input elaboration are likely to provide L2 learners with better input and opportunities to learn more lexical items incidentally.

EFFECTS OF DIFFERENT TYPES OF AUDITORY INPUT ON INCIDENTAL VOCABULARY LEARNING BY L2 JAPANESE SPEAKERS

by

Kyoko Kobayashi Hillman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Steven J. Ross, Chair
Professor Kira Gor
Professor Shoko Hamano
Professor Hong Jiao
Professor Jared Linck

To Yasuaki Kobayashi

Acknowledgements

This project was extremely rewarding for my professional career as both an SLA researcher and a Japanese language instructor. Although my investigation focused on enhancing L2 learning, I also learned fundamental components of SLA and its history, and how to conduct quality research.

It was an absolute honor to be able to work with Professor Mike Long and Professor Steve Ross on this project. I know there must be many SLA researchers and graduate students around the world who envy my position. I never imagined having such a privilege. Also, I would like to express my deep gratitude to my other committee members, Professor Shoko Hamano, Professor Jared Linck, Professor Hong Jiao, and Professor Kira Gor, who provided me with invaluable feedback and unfeigned encouragement. Furthermore, I would like to express my great appreciation to Professor Nan Jiang, Professor Robert DeKeyser, Professor Cathy Doughty, and Professor Lauretta Clough for their consistent support and leadership in our program. I will never forget their readiness and willingness to provide me with swift feedback no matter how busy they were.

In addition, I would like to extend my thanks to my Japanese friends who provided me with their written materials for the study, especially Ms. Mika Ebisawa, Mr. Akira Kawasaki, Professor Teruaki Enoto, and Professor Takao Okazawa. In particular, Mr. Teppei Kinami deserves special thanks for his assistance, spending hours in the lab to record all the talks. Rev. and Mrs. Himei in Tokyo also have my heartfelt thanks for their great help and hospitality when I was collecting data in Tokyo. Furthermore, I am thankful to the participants in the study and all of the people who helped me with my data collection.

Beyond this, I owe thanks to many friends/colleagues in the SLA program who inspired me and helped me with great encouragement in various ways. Without them, I would not have survived: Assma and Buthainah Al-Thowaini, Payman Vafaee, Yuichi Suzuki, Stephen O'Connell, Man Li, Hyojin Jeong, Nick B. Pandža, Fátima Montero, Jon Malone, Wei Yi, Kichan Park, Takehiro Iizuka, Ryo Maie, and Tetiana Tytko.

Furthermore, during the final stage of this process, I was able to begin working in the Department of Asian Studies at the University of British Columbia. The director, Professor Rebecca Chau, and my colleagues in the Japanese language program were very supportive, which made me feel extremely fortunate. Also, I must gratefully acknowledge Laurie Durand for her thorough assistance with editing my manuscripts. I would also like to thank Professor Junko Mori for her constructive advice whenever I contacted her. Patricia Novinski deserves a special acknowledgement for her consistent support and faith in me.

Lastly, I am very thankful to all of my family members and relatives, especially my parents, Masako and Norichika Kobayashi. My deep gratitude also goes to my husband, Ian, and Traffic and Jiro. I cannot thank you enough for your understanding and strong support. I know that all of you graciously spared me from my family duties during my Ph.D. studies. I feel very sorry that my father could not see me graduate, and that my brother, Yasuaki, passed away in my second year. I decided to go back to school for him, and it is to him that I dedicate this dissertation.

T.G.B.T.G.

Table of Contents

List of Tables

List of Figures

List of Abbreviations (alphabetical order)

CQ	Comprehension questions
EFL	English as a foreign language
ESL	English as a second language
FRT	Form-recognition vocabulary post-test
IVL	Incidental vocabulary learning
L1	A learner’s first language
L2	A learner’s second or foreign language
MDT	Form-meaning-recognition definition post-test
MST	Form-meaning-recognition sentence post-test
RHM	Revised hierarchical model
SLA	Second language acquisition
TW	Target word
WM	Working memory

Chapter 1 Introduction

Incidental vocabulary learning (IVL) is one of the primary research topics in second language acquisition (SLA; González-Fernández & Schmitt, 2017; Hulstijn, 2003; Laufer & Hulstijn, 2001; Laufer & Nation, 2012; Long, 2017). In IVL, vocabulary learning is a by-product of a primary activity to which the learner's attention is drawn. In other words, when IVL occurs, the learner's actual intention of learning must be focused elsewhere, such as on listening to news on the radio, because incidental learning excludes the intention to learn language, such as lexical items and grammar (Schmidt, 2010).

IVL studies on L1 children have reported vocabulary gains through reading and listening (e.g., Elley, 1989; Nagy, Herman, & Anderson, 1985). Whether adults learn novel L1 lexical items incidentally from only a few exposures is debatable (Hulme, Barsky, & Rodd, 2018). In the case of adult L2 learners, some evidence of IVL has been reported, but findings on the effects of instruction in IVL are mixed. The most investigated area is exposure frequency. Although findings generally show that more frequent exposure increases gains, Uchihara, Webb, and Yanagisawa (2019) argued that a plateau effect takes place after a certain number of exposures (Elgort, Brysbaert, Stevens, & Van Assche, 2018). Moreover, L2 IVL studies have assessed vocabulary gain through multiple types of post-tests, and different assessment types may lead to different results (e.g., Chen & Truscott, 2010; Malone, 2018; Van Zeeland & Schmitt, 2013; Webb, 2005, 2007a, 2007b).

In fact, the research findings in the IVL literature are difficult to synthesize due to methodological variations of at least five kinds: (a) definitions of IVL differ between online measurements, such as gaze duration and total reading time in eye-tracking, and offline measurements, such as vocabulary recognition and production; (b) measurement instruments differ, and often have problems with internal validity and sensitivity; (c) there is considerable variability in materials, including target words and texts; (d) ways of establishing incidental conditions differ; and (e) very few studies account for the influence of individual difference factors, such as working memory.

Compared to intentional vocabulary learning, vocabulary gains in IVL have been reported to be small. Therefore, questions regarding IVL, such as how to draw learner attention to target words and how to facilitate deeper lexical processing, have been raised as critical issues. Some studies have investigated the effects of input enhancement and/or input modification to tackle these issues. Findings in the input enhancement literature suggest that effects are limited to word form recognition (Issa & Morgan-Short, 2018; Lee & Huang, 2008; Leow & Martin, 2017). On the other hand, despite being the focus of fewer studies, the effects of input modification on IVL have been confirmed (Kim, 2006; O'Donnell, 2009; Toya, 1992; Urano, 2000; Vidal, 2011). Input modification (input simplification and elaboration) has also been reported to have beneficial effects on reading comprehension (Oh, 2001; Yano, Long, & Ross, 1994). And while many input modification studies have investigated IVL through reading using texts rigidly controlled for word frequency (Godfroid, Boers, & Housen, 2013; Hatami, 2017; Nation, 2001, 2013; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003; Webb, 2007a), their findings are not necessarily promising for pedagogical implementations.

To further investigate the relationship between IVL and input modification, this study focuses on IVL through auditory input modification. Specifically, using four different types of auditory input (genuine, simplified, elaborated, and modified elaborated), the study examines: (a) whether input modification facilitates listening comprehension, and (b) whether input modification enhances IVL. The outcome variables were: (a) three types of comprehension questions, and (b) three types of unannounced vocabulary post-tests. Participants were advanced-level L1-Chinese speakers of L2 Japanese residing in Japan. Due to the input being auditory, results were assumed to be associated with working memory (WM); therefore, the participants' working memory capacities were also assessed. In addition, participants completed an L2 proficiency test, a background questionnaire, and a debriefing questionnaire.

Chapter 2 reviews the relevant literature, beginning with the L2 IVL literature. The chapter covers theoretical frameworks concerning underlying mechanisms, and then moves onto input modification, listening comprehension, and WM in L2 listening. Chapter 3 presents the research questions and discusses the expected results of this study. Chapter 4 describes the methodology, including participants, materials, instruments, and procedure. Chapter 5 reports the results of the analyses. Chapter 6 discusses the results and findings. The study concludes with a summary and directions for future research in Chapter 7.

Chapter 2 Literature Review

2.1 Incidental vocabulary learning

Despite some pedagogical approaches that recommend effective vocabulary instruction (Barcroft, 2016; Nation, 2001; Webb & Nation, 2017), the teaching and learning of vocabulary has been under-researched until recently, especially in comparison to grammar instruction. There may be several reasons for this research gap, including a sense of vocabulary as secondary to grammar; we still lack influential theoretical frameworks for vocabulary learning. Another reason may be a conventional expectation for language learners to study and learn vocabulary on their own, using the various tools available, such as dictionaries and flash cards.

Nation (2001, 2013) provided a comprehensive overview of relevant empirical studies, along with a synthesis of information directly applicable to the teaching and learning of L2 vocabulary. He recommended setting specific goals for L2 learners and teachers based on word frequency, rather than focusing on the size of native speakers' vocabulary. The pedagogical approach Nation proposed consists of four major strands: (a) learning from comprehensible meaning-focused input through reading and listening, (b) learning from meaning-focused output through writing and speaking, (c) language-focused learning (i.e., form-focused instruction), and (d) developing fluency. As the third strand suggests, he supported deliberate teaching and learning.

In Nation's view, the distinction between "intentional" and "incidental" learning is not as important as "the quality of the mental processing that takes place during learning" (2013, p. 349). In outlining specific instructional goals for facilitating the necessary quality of mental processing, Nation (2001, 2013) argued that the ratio of unknown words in a reading text should be a maximum of one for every 20 running words (i.e., counting tokens); in other words, learners should be familiar with at least 95% of the words in a text for adequate comprehension. For example, to understand newspapers, a non-native speaker of English needs knowledge of 4,000 word families to achieve 95% coverage, and 6,000 word families for 98% coverage. Under such conditions, Nation claimed, IVL is likely to occur, as a learner will simply infer the meaning of unknown words from the immediate context. Therefore, he suggested, the context and the type of context can facilitate or hinder the learning of unknown words. When contextual information is available, learners might be able to infer the meaning of a word even at the first encounter. Nation further argued that vocabulary-learning strategies can help learners develop vocabulary knowledge through inferencing.

While the use of authentic teaching materials is strongly encouraged in current foreign language education practices (see, e.g., Breiner-Sanders, Swender, & American Council on the Teaching of Foreign Languages, 1999), Nation's (2001, 2013) approach demonstrates an attempt to discover the types of texts most suitable for learning L2 vocabulary from reading. One weakness of this approach is the difficulty of controlling the coverage of known words in a text for learners at advanced L2 levels. It is not clear whether this pedagogical approach is designed to apply to more advanced learners.

However, modified texts, including word-frequency-based texts, could certainly be more effective for all L2 learners than authentic texts. Modified texts are texts that are altered for language learning by manipulating lexical items, linguistic features, and/or organization in such a way that L2 learners can understand the content. Whether modified texts are better for L2 learners, and what kind and type of modified texts enhance L2 learning as input, remain unresolved issues, although they have often been explored. IVL studies looking into these and other questions, such as number of exposures and input modes, will be reviewed in the next section.

2.1.1 Number of exposures

Most IVL studies have been conducted through reading (e.g., Nagy et al., 1985; Pellicer-Sánchez, 2015; Waring & Takaki, 2003; Webb, 2007a, 2007b, 2008). For instance, Nagy et al. (1985) found that 57 eighth-grade children showed small vocabulary gains from reading. These studies generally support at least three conclusions: (a) IVL occurs while reading in L1 and L2; (b) vocabulary gains from reading are smaller than in intentional vocabulary learning conditions; and (c) the effects of IVL should be explored, because intentional learning alone takes too long to meet most L2 learners' vocabulary learning goals.

In this line of research, one central focus has been the number of input exposures needed to facilitate new lexical learning: in other words, the effects of repetition. Findings generally suggest that a higher number of exposures to new items in reading results in higher scores on unannounced vocabulary tests of both form and meaning. This effect appears to be the same in both L1 and L2 vocabulary learning through reading (e.g., Godfroid et al., 2017; Hulme et al., 2018; Pellicer-Sánchez, 2015; Webb, 2007a).

Waring and Takaki (2003) investigated IVL through reading, assessing the retention and decay of newly learned words over time, the number of input exposures needed for meaningful learning, and types of test formats for vocabulary knowledge. Participants were 15 EFL students at an intermediate proficiency level. Using available graded readers at the lowest level, they created a reading passage, in which the known/unknown word coverage rate was 96.2%. Twenty-five target words (TWs) were chosen from the text and divided into five groups depending on the number of occurrences in the text (1, 4–5, 8–10, 13–14, and 15–18 occurrences). The TWs were substituted with non-words, which were constructed by changing the spelling of the original TWs in the text, to avoid the influence of pre-existing lexical knowledge. Gains were assessed through three post-tests: a form-recognition test, a multiple-choice meaning-recognition test, and a meaning translation test into the L1. The post-tests were administered three times: immediately, seven to 10 days later, and three months later. The study's definition of learning was narrowly framed, but considered the scores on the translation test to be valid for assessing learning. Results indicated that participants learned a relatively small number of TWs through reading; scores on all three test types were not high even at the time of the immediate post-tests, and consistently decayed over time. There were, however, clear differences in the scores from the different test types, suggesting that each test type tapped into a different aspect of lexical knowledge. On the immediate post-test, the form-recognition test's mean score was the highest (61.2%), and the meaning translation test's mean score the lowest (18.4%). On the three-month delayed test, the mean score of the form-recognition test decreased by 33.6%, while the mean score on the translation test dropped by 3.6%. Rank order of scores on the tests (form > multiple-choice > translation) remained the same over time. The overall results suggest that retention of form is stronger than retention of meaning, even though the graded reader provided participants with contextual information regarding meaning. After three months, the decay rate of form-recognition and meaning-recognition was over 40%.

The results indicated that an effective number of word occurrences for learning was at least eight times for a 50% chance of recognition after three months, and more than 18 times for long-term retention of meaning by translation. The study was based on the assumption that the known word coverage (over 96%) was high enough to facilitate IVL; however, the findings suggested only small effects, and the authors interpreted the results as constituting weak evidence of IVL. Waring and Takaki suggested the possibility that adding text modifications, such as highlighting to draw attention to TWs, could lead to better IVL outcomes.

Webb (2007a) also investigated exposure frequency in IVL while reading. Although previous studies had investigated the necessary number of exposures, their suggestions varied widely: more than 18 times (Waring & Takaki, 2003); 10 times (Saragi, Nation, & Meister, 1978); eight times (Horst, Parsons, & Bryan, 2011; and six times (Rott, 1999). Webb used four conditions of exposure frequency in reading passages (1, 3, 5, and 10 exposures). He suggested that the variety of previous results was caused by insufficient and unreliable instruments, and therefore devised 10 different instruments to measure vocabulary gains, focusing on different dimensions of lexical learning. Results confirmed repetition effects, but also confirmed learning gains from three exposures on all 10 immediate post-tests. Gains were mostly limited to receptive knowledge with one exposure, while gains from three exposures included productive knowledge. Chen and Truscott (2010) supported Webb's findings, reporting that seven exposures was sufficient to improve production, based on the results of immediate and delayed post-tests. Van Zeeland and Schmitt (2013) suggested 11 exposures were necessary for IVL from reading, but fewer were needed for IVL from listening. Godfroid et al.'s (2017) eye-tracking study indicated that different types of cognitive processes could be involved in IVL, depending on exposure frequency. They pointed out a slight increase in fixation times between the seventh and tenth exposures. They argued that this finding could reflect sentence-integration for semantic processing.

Pellicer-Sánchez (2015) also used eye-tracking to investigate L2 IVL through reading. The study examined lexical knowledge gained from reading, using both online and offline measures. Three types of offline vocabulary post-tests, as well as four types of online reading times measured through eye-tracking, were used. Twenty-three advanced English as a second language (ESL) students and 25 native speakers of English participated. The data obtained from L1 speakers were used as the baseline. Participants read a short story containing six non-word TWs embedded in the text eight times, along with six control words. One of the research questions asked how TW total reading times would change across exposures. Online measures showed that total reading times for the TWs decreased after the third or fourth encounter. After eight exposures, TW reading times overlapped with those for familiar words. Moreover, results of the online measures showed a clear relationship between the words L2 participants could recall and total reading times for the words: L2 participants' reading times for the words that were better recalled were significantly longer than those for words that they could not recall. These results suggest that participants could recall the words that they spent more time on when reading.

Pellicer-Sánchez (2015) interpreted these longer reading times as reflecting attempts to infer the meaning of the unknown TWs. Results of offline measures revealed no statistically significant difference in vocabulary gains between L2 and L1 participants.

Mean scores for both groups on all of the post-tests were high, although the three different types of post-test showed slightly different results. The gains found in this study were larger than those found in previous studies. The L2 participants in this study recognized over 85% of the six TWs on the immediate and the two-week delayed post-tests. They were able to recognize the meaning of over 75% of the TWs on both post-tests. The scores of the meaning recall tests were not as high, with mean scores on the immediate post-test at 61% for the L2 group, dropping to 55% on the delayed post-test. The overall results, however, indicate retention even after two weeks. Because all participants encountered the relatively small number of TWs eight times during the treatment, frequency effects became apparent.

Unsurprisingly, research shows that greater exposure is preferable for steady lexical development during IVL. However, variability in materials (including TWs and contextual information), along with un-accounted for individual differences among participants, make it difficult to draw conclusions regarding the extent of the influence of exposure frequency in IVL while reading. For instance, the number of TWs in treatment and exposure conditions could influence outcomes. Waring and Takaki (2003) study used 25 TWs in one treatment, with variable frequency of occurrence. The length of a short story in their study was 5,872 words. On the other hand, Pellicer-Sánchez (2015) used six TWs, with frequency of occurrence rigidly controlled at eight. The length of a story in her study was 2,300 words. Although repetition effects are confirmed as robust (e.g., Uchihara et al., 2019), the minimum number of necessary exposures remains unclear, and should be further explored in conjunction with other factors. The effects of several factors, including text variability (e.g., brevity, complexity), number of TWs, type of TWs, and frequency of exposure are likely to be interconnected in IVL.

2.1.2 Multimodality

Vidal (2011) explored IVL in the academic context by comparing learning gains in reading and listening. Participants were L1-Spanish learners of L2 English, and their intact classes were assigned to three different treatments: (a) reading three academic texts, (b) watching three lectures, and (c) receiving no input (control). Thirty-six TWs were classified as technical, academic, or low frequency, in accordance with the categorization by Nation (2001). Words were presented with implicit lexical elaboration, explicit lexical elaboration, or no elaboration. Implicit elaboration included paraphrasing and synonyms, and explicit elaboration included formal and semi-formal definitions, as well as descriptions, naming, and questioning statements. Frequency of occurrence was controlled, across passages, from one to six. While the TWs' meanings were intended to be impossible to deduce in isolation, they were classified into four categories based on the predictability of their meaning (similar to Spanish, morphologically predictable, deceptively transparent, unpredictable). The TWs were selected considering crosslinguistic similarities in orthography, morphology, and semantics, as well as frequency in both Spanish and English; chosen TWs were low in frequency in both languages. A pre-test, which took place approximately four weeks prior to the treatment sessions, was administered to assess participants' knowledge of the TWs. The treatment was divided into three sessions. Results were analyzed in a 3 x 3 ANCOVA (with tests, groups, and L2 proficiency as covariates), which showed a significant three-way interaction effect. This effect was due to the variety of L2 proficiency levels among the participants. One of the findings was that overall vocabulary gains were greater from reading academic texts than from listening to lectures. However, participants whose proficiency level was higher showed vocabulary gains from both reading and listening. Elaboration types worked differently in both groups as well. Both explicit and implicit lexical elaboration showed effects in listening, but only explicit elaboration was effective during reading. Results regarding the relationship between vocabulary gains and TW characteristics showed different predictors between regression models for reading and listening. The best predictor for reading was frequency of occurrence, and that for listening was predictability from form (e.g., false cognates). Vidal suggested that phonological memory plays an important role in IVL while listening. Critically, she reported robust effects of explicit elaboration on links between form and meaning in both reading and listening.

Brown, Waring, and Donkaewbua (2008) compared vocabulary learning through three different input modes: reading only, reading-while-listening, and listening only. Participants were 35 undergraduate students in Japan, at pre-intermediate levels of English proficiency. They were randomly assigned to the three input mode groups, and the materials were three stories from 400-headword graded readers at a high beginner level. The total of 28 TWs was embedded in each story. The TWs were altered spellings of actual words. Frequency of occurrence was controlled in four groupings (2–3, 7–9, 10–13, and 15–20 exposures) for each seven TWs (7 TWs x 4 groups). The proportion of familiar words, or coverage rate, in each passage was approximately 95%. All three groups were counterbalanced for the three treatments, and participants completed two types of meaning-based vocabulary test: L1 translation and multiple choice.

Brown et al. (2008) reported better learning outcomes than those found by previous studies. Neither test type showed any significant difference in scores between the reading only and the reading-while-listening groups. However, the scores of the listening only group were significantly lower than those of the other two groups, especially on the immediate translation test. The study concluded that learning through listening alone was too difficult for the participants, even with frequent exposure (i.e., the maximum of 15–20) to the new lexical items. However, it is worth noting that one listening passage took approximately 60 minutes, without a break. In addition, participants were not allowed to take notes while listening. Therefore, other factors related to the setting and the learners may have contributed to making the listening-only mode particularly difficult for IVL.

Hatami (2017) compared vocabulary gains in IVL through reading and through listening. One hundred and thirty-nine pre-intermediate EFL students were randomly assigned to reading, listening, and control groups. Participants read or listened to a graded reader modified for the first 1000-word frequency level. Vocabulary possibly unknown to participants was altered to easier lexical items. Sixteen TWs were selected, based on their part of speech and frequency of occurrence in the original text. Subsequently, TWs were altered to English-like non-words, assuming the original TWs were known to the participants. The reading/listening treatment was administered without comprehension questions. The control group did not receive a treatment. Dependent variables were measured using six unannounced vocabulary post-tests: (a) recognition of spoken form,

(b) recognition of written form, (c) recall of meaning, (d) recognition of part of speech, (e) recognition of syntagmatic association, and (f) recognition of meaning (Chen & Truscott, 2010; Van Zeeland & Schmitt, 2013; Webb, 2005). The same set of vocabulary tests was administered three weeks later as delayed post-tests. Results of the reading group on the immediate post-tests showed the largest gains on written form (75% of TWs), followed by part of speech (55%), meaning recognition (54%), spoken form (53%), syntagmatic association (49%), and meaning recall (14%). On the other hand, the results of the listening group showed the largest gains on spoken form (50%), followed by written form (45%), part of speech (37.3%), meaning recognition (37%), syntagmatic association (25%), and meaning recall (6%). The patterns for the larger gains were clearly different, and strongly associated with the input mode. Except for recognition of spoken form, the tests were administered in the written mode. Therefore, as Hatami pointed out, the reading group would have benefited more from the testing mode than the listening group, which was a limitation of the study. Another limitation was that the results revealed testing effects on all of the recognition tests. Therefore, retention after three weeks was analyzed only for meaning recognition and recall. Scores for the reading group on the meaning recall immediate post-test were significantly higher than those for the listening group, and the significant difference between the groups remained the same on the delayed post-test. On the meaning recognition post-tests, the reading group scored significantly higher than the listening group on the immediate test, but three weeks later, there was no statistically significant difference between the two groups. The listening group scores on the meaning recall and recognition tests remained almost the same between the immediate and delayed tests.

Malone (2018) investigated the effects of exposure frequency and aural enhancement (AE) in reading. Eighty ESL students were randomly assigned to four treatment groups: (a) two exposures with AE, (b) two exposures without AE (visual reading alone), (c) four exposures with AE, and (d) four exposures without AE. Participants read four stories that contained 32 very low-frequency (beyond the 25,000 band range) TWs. Two types of unannounced post-tests were administered: (a) a form-recognition test, in which participants circled all the words they recognized from the reading passages; and (b) a multiple-choice form-meaning connection test, in which participants circled the correct semantic category for each TW. There were 32 distractors in the form-recognition test and two for each item in the form-meaning connection test. Comprehension questions were included to ensure a focus on meaning, and thus create an incidental learning condition. An L2 proficiency measure and three WM measures were employed as covariates. Confirmatory ANOVA analyses were conducted using estimated marginal means and standard deviations of the groups. Results of the analyses showed group differences indicating separate effects of exposure frequency and input mode. Frequency effects on the form-recognition test were greater than AE effects, and AE effects were confirmed at two exposures. On the other hand, both frequency and AE influenced outcomes on the form-meaning test. Malone suggested that AE helped semantic processing during the reading of visual input. This finding supports previous findings in the literature suggesting that L2 vocabulary knowledge consists of different components (e.g., Pellicer-Sánchez, 2015; Webb, 2005). The study’s regression analyses investigated the relationship between form-recognition and WM in each group, and the results showed that the AE groups required more WM than the visual reading groups.

This finding suggests that aural input during visual reading placed an additional burden on the participants' WM. Moreover, the results indicated that the relationship between form-meaning connection and WM was not strong. Malone raised a question as to whether IVL is less likely to occur through listening than through reading because IVL while listening may depend more purely on WM resources.

In a study based on the output hypothesis (Swain, 1995), Nguyen and Boers (2019) reported on vocabulary uptake in conjunction with an output task after audio-visual input. Participants were 32 upper intermediate EFL students in two intact classes, with both groups watching a TED talk once. Afterwards, one group completed a summary task, and then both groups watched the video again and took post-tests on 18 TWs from the video. Results showed that the summary task group made significantly higher gains on both post-tests, with large effect sizes for both uptake (d = 0.81) and retention (d = 0.90). The mean scores for the immediate post-test were 45% for the summary group and 26% for the control group. These results show that the summary task enhanced IVL by approximately 20%, although the control group also exhibited gains from two viewings of the video.

The emphasis on the output hypothesis in Nguyen and Boers (2019) is unique in this line of research. The task they used is ecologically valid, at least in a classroom setting. An important caveat is that the students were allowed to take notes and review them while answering comprehension questions. This procedure appears to be closer to an intentional learning condition, because the students were forewarned that they would be asked to sum up the video later. In this situation, the students may well have thought about key words to remember, so that they could reuse them during the summary task.

The control group may have been in a genuine incidental learning condition, because they did not perform the summary task. Nguyen and Boers claimed to have established an incidental learning condition by not announcing the vocabulary post-tests. However, it might have been more informative to include another step, such as debriefing at the end of the procedure, to verify whether any learning that took place was indeed incidental.

Interestingly, Nguyen and Boers (2019) also mentioned the effects of visual cues and input modification (p. 22). During the TED talk, one TW happened to be accompanied by visual and verbal cues, and another by lexical elaboration. The support of these cues, which occurred naturally during the talk, appeared to affect scores on those two words, suggesting facilitation effects on vocabulary learning. Comparable findings have been reported in other studies (e.g., Elley, 1989), which suggests that how these cues support IVL should be explored further.

2.1.3 Auditory input only

Few studies have explored IVL through auditory input alone. Elley (1989) investigated whether children learn L1 vocabulary incidentally while listening to stories read aloud by teachers. Elley also investigated the effects of teacher explanations of novel lexical items compared with reading alone (i.e., with no explanations). One hundred and twenty-seven eighth-grade children participated in her second experiment in two treatment groups, along with 51 children in a control group. A pre-test, containing 36 TWs from the experimental texts, along with five control words, was administered to test pre-existing vocabulary knowledge. The words in the pre-test were presented in different contexts from those in the experimental texts. In the treatment, the children listened to two stories read aloud by different teachers three times in a given week. One group listened to the stories with teacher explanations for the TWs. The teachers, following provided guidelines, explained the TWs in one of three ways: (a) using a synonymous phrase, (b) role-playing or acting out, and (c) pointing to a picture. The other group did not receive TW explanations while listening to the stories. The control group did not listen to the stories at all. Unannounced post-tests for TWs were administered to all groups seven days after the end of the treatment, and again three months later. The test was multiple choice; half of the choices were pictures and the other half were synonyms of the TWs.

Results showed that the non-explanation group's mean score increased by 14.8% compared to their pre-test mean score. This was comparable to the outcomes of her first experiment, in which the treatment was listening to a story without TW explanations. The group with teacher explanations showed much larger gains, with a 39.9% mean increase. No testing effects were observed in the results from the control group or the five control items in the test. The results of the post-test administered three weeks later did not show any substantial decay for either treatment group, indicating only a 2–3% decline from the first post-test results. Therefore, these results showed that IVL occurred through listening to the stories, and that vocabulary gains were greater when they received explanations of the meaning of TWs. In addition, Elley reported that the amount of gains varied depending on the book. Therefore, Elley noted, the story itself also likely affected the gains. This trend was also pointed out by Nagy et al. (1985).

Additionally, Elley (1989) identified six word-related factors: the number of occurrences in the text; the number of pictorial occurrences (i.e., TWs shown with pictures); the helpfulness of the verbal meaning cues; the importance of the word to the development of the plot; the vividness of the word; and the likely familiarity of the concept for the participant population. Whether these factors had a systematic relationship with vocabulary gains was examined. The first two factors were found by counting the numbers of TWs in the stories. For the other factors, mean ratings by the six participating teachers were used. Results showed that all six variables were statistically significantly correlated with vocabulary gains. The highest correlations were found with the number of occurrences in the text (.60) and the number of pictorial occurrences (.55). Elley used pictures not only during the treatment, but also in the post-test. Therefore, the effects of pictures on IVL showed more strongly than the effects of other factors. Similarly, Bisson, Van Heuven, Cocklin, and Tunney (2015) reported the effect of pictorial cues as the best for IVL, while reporting that auditory input alone resulted in the lowest vocabulary gains.

Van Zeeland and Schmitt (2013) conducted a study of IVL through listening. They used four listening passages, such as TV interviews, available from the Internet. These passages contained lexical items within the 2,000-word frequency band. They used 24 non-word TWs (8 nouns, 8 verbs, and 8 adjectives) in the passages. Target words were selected from the original texts. Because the participants knew the target words, non-words were generated with an online non-word database, and the original words were replaced with the non-words (e.g., *grath* for "house," *to cluss* for "to understand"). Thirty ESL graduate students whose English levels were at the intermediate-high or advanced level listened to the passages for 20 minutes, and were exposed to the TWs at four different frequencies (3, 7, 11, or 15 times). Three types of post-tests were administered to assess three dimensions of vocabulary knowledge: (a) a multiple-choice form-recognition test with four options, (b) a grammar-recognition test in which participants selected the part of speech of the TWs, and (c) a meaning-recall test where the participants wrote down anything they knew about the TWs.

Results from immediate and two-week delayed post-tests showed less learning of the meanings of the TWs than of their forms and grammar (i.e., the part of speech). Overall learning gains were found to be small, and varied depending on word concreteness. Results on the delayed post-tests, however, indicated different patterns of retention. Two weeks later, scores on the form-recognition and grammar tests had significantly decreased. The decay was considerable, with a decrease of approximately 20%. In contrast, scores on the meaning-recall tests showed no statistically significant difference between the immediate and delayed post-tests. These findings align with those reported by Hatami (2017), which also indicated no significant difference between immediate and delayed post-tests in a listening group's mean scores on meaning recall and recognition tests. The results of Van Zeeland and Schmitt's delayed post-test showed that exposure frequency did not affect the retention of the different dimensions of vocabulary knowledge, except for significantly higher scores from 11 exposures. The scores of form and grammar recognition after 11 exposures on both immediate and delayed post-tests were the highest.

Van Zeeland and Schmitt (2013) pointed out that smaller learning gains from listening than reading are consistently reported in the literature. They suggested three reasons for smaller gains from listening: (a) the need for faster processing, (b) the difficulty of segmenting spoken input, and (c) the difficulty of spotting new words. They also summarized the effects of frequency (i.e., occurrences/exposures) found in the literature on IVL from either listening or reading passages, reporting that three times was the minimum, and 11–15 times the upper threshold, for measurable effects. Their reason for naming three exposures as the minimum is unclear, as Malone (2018) reported vocabulary gains after two exposures. In any case, however, they emphasized that frequency of occurrence is only one factor influencing IVL, supporting the suggestion that other factors affecting IVL should be further examined.

2.1.4 Contextual information

The definitions of “context” and the types of passage used as context vary in the L2 IVL literature. Experimental materials have employed sentences, settings, paragraphs, passages, and books as contexts for TWs (Elgort, Brysbaert, Stevens, & Van Assch, 2018; Godfroid et al., 2013, 2017; Van den Broek, Takashima, Segers, & Verhoeven, 2018; Vidal, 2011).

Webb (2008) speculated that the lack of clarity about IVL in the field has been caused not only by mixed results on the sufficient frequency of occurrences/exposures, but also by the variety of contexts that have been used. To investigate the effects of context, Webb constructed 30 sentence-long passages of two types. The types differed with regard to the level of detail pertaining to the meanings of the 10 TWs. The degree of inferable information was rated for each TW by native speakers of English. The experimental group read the more informative passages, while the control group read the less informative passages. Each TW appeared three times in both conditions. Afterwards, the students took four unannounced vocabulary tests. The post-tests focused on both recall and recognition of forms and meanings. The recall tests required the students to write the TWs and to translate them into their L1. The recognition tests were multiple-choice tests with four word options. There were only 10 TWs, and the post-test did not provide contextual information. Results showed higher scores for the experimental group than for the comparison group. While the overall scores were significantly different between the groups, the scores on each separate post-test were not.

Another study by Webb (2007b) also showed the effects of context, this time at the sentence level. The original purpose of the study was to investigate the effects of synonyms when learning novel lexical items incidentally. Participants were EFL university students in Japan, and treatment conditions included TWs with or without synonyms in word pairs (TW and L1 translation word) or in sentences (where the TW was glossed and accompanied by an L1 translation). Results did not show a statistically significant difference between conditions with and without synonyms. All presentations of TWs included L1 translation words. Thus, participants might have relied on the translations more than the synonyms. However, the results showed that almost all scores in the recognition post-tests of the group exposed to the TWs in sentences were higher than those of the group exposed to the TWs in word pairs (with no context). Although Webb’s original research questions did not address the different types of presentation, the results clearly favored the sentence-level context for TW recognition.

The IVL studies conducted by Webb (2007a, 2007b, 2008) used nonsense words. His justification for this methodological choice was to avoid the possibility of pre-existing knowledge of the TWs and the difficulty of including low-frequency words in graded readers, because graded readers are constructed within a restricted range of frequency bands. The 10 nonsense TWs in these studies consisted of six nouns and four verbs. They replaced high-frequency words that were likely to be known to the students (e.g., *hospital*, *lunch*, *to visit*). Due to the strict control over the vocabulary used in the contexts, the only words unknown to the participants were the nonsense TWs. Under this condition, when reading the passages, the saliency of the TWs should have been high. However, the participants could easily guess the original high-frequency words to understand the content of the reading passages. There might have been no deeper processing due to high inferrability, but this high inferrability may have inhibited participants' meaning learning of the experimental nonsense words. The study could have been more informative if comprehension questions had been asked of the participants, as their responses might have revealed relationships between their comprehension of contextual information, inference of the meaning of the TWs, and actual learning of the TWs.

Furthermore, due to the high saliency of the TWs' presentation, it is unclear whether an incidental learning condition was established. In one of these studies (Webb, 2007b), both groups saw the TWs with L1 translation words. One group was told to study the word pairs (the TWs and the L1 translation) for eight minutes. The other group read whole sentences, but the TWs were glossed and presented with the L1 translation words. This treatment appears to be an intensive word-recall activity. Therefore, the students were likely to remember the TWs intentionally even without knowing the true purpose for learning them (i.e., taking the vocabulary tests, which were unannounced).

As Nation (2001, 2013) observed, the line between incidental and intentional learning appears to be a fine one. Controlling learner attention is the key to developing tasks for receptive skills to establish incidental conditions. Some studies have overstepped this fine line.

Elgort et al. (2018) conducted an eye-movement study to investigate the process of IVL in academic reading. Their focus was how TWs that are low in frequency become familiar to participants across exposures and how well the words are integrated into the participants' lexicon. The reading passages used in the study were an introduction and a chapter from an academic book that included vocabulary at the level of 6,000-word frequency families. Analysis of gaze duration found decreases at the fifth and seventh TW encounters, as well as a change of total reading times at the tenth encounter. After determining that the eighth encounter was the cutoff point, they found a statistically significant difference in gaze durations between TWs and controlled words in reading, and the mitigation of the reading time by the eighth occurrence in the text, using mixed-effects modeling. These results suggested that, with increasing encounters, the TWs became closer to the controlled words in terms of lexical access and word-to-text integration. The results also support the finding of eight encounters as the threshold for processing time during IVL in Pellicer-Sánchez (2015). However, the study by Elgort et al. also suggested that IVL in reading genuine academic texts is a slow and incremental development. It also used a meaning generation task (i.e., meaning recall task) as an offline measure. The mean score on the task was 34%. This score indicates overall low accuracy on meaning recall, suggesting that the participants' TW knowledge was not sufficient to capture the meaning of newly learned TWs, a finding that aligns with others in the literature. The study also suggests "a possible dissociation between the development of knowledge supporting word understanding during reading and knowledge allowing participants to explicitly formulate the meanings of words learned in contextual word learning" (pp. 360–361) due to the lack of a reliable relationship between the eye-movement measures on the reading post-test for vocabulary knowledge and the written meaning-generation task scores.

Taking these results together, word learning from context appears to be slow and incremental (Elgort et al., 2018; Godfroid et al., 2013, 2017; González-Fernández & Schmitt, 2017; Webb, 2008). A few studies have explored the use of contextual information in vocabulary learning under an intentional learning condition. Van den Broek et al. (2018) argued that contextual information did not enhance word retention, and that word retrieval was more effective. Barcroft (2015) also explored possible effects of new word retrieval in reading a story. Both studies led participants in a retrieval task during a reading phase, so that the participants could use contextual information to learn novel lexical items. The quality and quantity of contextual information in the two studies differed. Van den Broek et al. used two sentences to enhance vocabulary item retrieval through richer contextual information, while the control condition used only one sentence. Barcroft emphasized the importance of discourse-level contextual information, asserting that discourse-level information could enhance IVL by providing multiple aspects of target lexical items, including the natural language environment in which the items tend to occur. Although the treatment in both studies remained as unobtrusive as possible without providing explicit instruction for vocabulary learning, the procedures did not create a clear setting for IVL due to the use of an explicit form-focused task (i.e., a fill-in-the blank task in the text to retrieve words).

In sum, there is a host of factors affecting IVL. The most explored to this point has been the frequency of word occurrences, or exposures. Input modality has also been investigated. The effects of exposure frequency on IVL could differ depending on input modality (Hatami, 2017; Malone, 2018; Van Zeeland & Schmitt, 2013; Vidal, 2011). Most of the research pertaining to IVL has focused on reading, rather than on listening. Possible reasons for this are the difficulties in investigating cognitive processing during listening and in measuring vocabulary gains that are expected to be smaller than gains during reading tasks. The rate of reading and/or listening could also affect IVL. Furthermore, given that both reading and listening are receptive, individual differences, such as working memory and L2 proficiency, are considered to play a role in IVL.

Treatment procedure affects IVL as well. In a meta-analysis of IVL based on spoken input, De Vos, Schriefers, Nivard, and Lemhofer (2018) included whether post-tests were announced prior to the treatment in their guidelines for selecting studies. To make vocabulary learning incidental, IVL studies are required to have surprise post-tests. However, a few studies appeared to justify the condition only because the post-tests were unannounced. The treatments in the studies did not appear to keep the true purpose of vocabulary learning opaque, and when conducted with intact classes, students could have easily sensed what they would be expected to do after the task or activity. As discussed earlier, the line between incidental and intentional can easily be crossed, depending on the instruction. The implications of IVL studies' findings may be strongly associated with the types of knowledge that develop through different learning conditions. IVL may result in implicit knowledge if the learner remains unaware of what is learned (Long, 2017). The learning process, however, may not continue to remain incidental if the learner's attention is redirected and the learning condition becomes intentional. Under intentional learning conditions, the resulting knowledge could be explicit (Malone, 2018). It is critically important to establish that incidental learning conditions are created. Therefore, the length of treatments, the inclusion of comprehension questions in meaning-focused treatments, and debriefing sessions to gather additional information regarding participants' attention and awareness during treatments must be considered when conducting IVL studies.

Many studies in the IVL literature use non-words as TWs. Waring and Takaki (2003) questioned whether it is more difficult for participants to learn non-words substituted for known words (e.g., *sind* for "snow"). The non-words could be salient enough to attract participants' attention in reading or listening passages that achieve lexical coverage of 95% or above (Nation, 2001, 2013). When Waring and Takaki interviewed their participants after the treatment, they found that the participants had been confused by the non-words. However, due to the high inferrability of the non-word TWs (i.e., participants could guess the original English words for the non-words from the passages), the reading process was not slowed down. Use of non-words eliminates the possibility that participants already know TWs and it is convenient for a delayed post-test, because participants will not encounter them after a treatment. However, it may create unnatural vocabulary learning conditions.

Webb (2008) pointed out a possible issue with the variety of contexts in the IVL literature. As discussed, this variety includes both quantity and quality. Some studies have used sentence contexts, such as Webb (2007b) study that showed the effects of sentence context on recognition in comparison to word pairs. Although a sentence can provide contextual information for a target word, the amount of information provided is inherently limited. It is questionable whether such limited contextual information could have a substantial impact on IVL. Other studies have employed paragraph-level context, and some have used authentic reading passages and books as contexts, emphasizing the importance of ecological validity. Vocabulary manipulation in the context has also varied. Graded readers and equivalent contexts have been used, so that the only words unknown to the participants were the TWs. Other studies have used academic lectures and book chapters. Overall, regardless of the contexts utilized in the treatment, vocabulary gains have been small. Such small reported gains could possibly have been caused by the use of discrete-point post-tests. Eye-tracking studies, however, confirm gradual and sluggish lexical development. Van den Broek et al. (2018) explored whether a newly learned word can be retrieved from memory or inferred from rich contextual clues. Results suggested that contextual information helped learners find the correct word meaning during practice, but also appeared to lead to low retention rates. In their discussion of the mixed results, Van den Broek et al. suggested the possibility that the context inhibited the establishment of a clear connection between the form and meaning of a novel word, which resulted in no retention. The context could be ambiguous for TWs, whereas it is unclear how lexical coverage of 95% and above helps IVL. Participants' familiarity with the subject matter could also influence IVL. Elgort et al. (2018) suggested that the provision of context or no context resulted in possible differences in lexical knowledge. Few studies have included comprehension questions for written or spoken input. Therefore, the overall picture of IVL remains unclear, and more studies are needed.

Furthermore, reliable instruments for vocabulary learning have not been used in IVL studies even though measurements of gained vocabulary knowledge are critical. Webb (2007a) devised 10 different instruments to assess vocabulary knowledge: (a) productive orthography, (b) receptive orthography, (c) receptive recall of meaning and form, (d) receptive recognition of meaning and form, (e) productive association, (f) receptive association, (g) productive syntax, (h) receptive syntax, (i) productive grammar, and (j) receptive grammar. Some other studies have adopted these instruments (Chen & Truscott, 2010; Hatami, 2017; Van Zeeland & Schmitt, 2013). Webb (2007b) also created a different set of instruments to assess orthography, syntagmatic association (the relationship between elements combined with each other; for example, *locomotive–station* or *locomotive–arrived*), paradigmatic association (the relationship between elements of the same category, i.e., elements that can be substituted for each other; for example, *locomotive–vehicle* or *locomotive–airplane*), grammatical functions, and meaning and form. His intention in creating these tests for word associations was to examine vocabulary knowledge with synonyms. Other studies have used conventional measures, such as the multiple-choice form and/or meaning recognition and recall tests. The recall tests in Webb (2007b) were productive; the form recall test required participants to spell out TWs, whereas the meaning recall test included L1 translation of TWs and writing down anything remembered about them. Elley (1989) used both pictures and synonyms for meaning recall. Most of the post-tests in these studies were discrete-point post-tests for TWs. Godfroid et al. (2013) used a meaning recognition test in the same sentence context as the passage. Participants were required to choose one appropriate TW out of 18 choices within 15 seconds. Few studies have employed delayed post-tests, possibly due both to expectations of low retention rates and to research constraints. The approach used by Webb is called a multi-component approach for vocabulary knowledge. As different types of immediate post-tests have indicated different results, the tests appear to tap into different dimensions of vocabulary knowledge. A construct of vocabulary knowledge gained from IVL specifically has not yet been examined.

Laufer and Hulstijn (2001) suggested a different perspective on IVL when they asked "whether the quality of exposures to new vocabulary during 'incidental' encounters can compensate for the relatively limited amount of exposure which is characteristic of learning a second language in a non-language-speaking environment" (p. 22). It is worth exploring whether there is a relationship between the quality of contextual information for a novel lexical item and uptake of it through IVL. If there is such a relationship, how contextual information can contribute to the uptake could be explored.

2.2 Underlying mechanisms

2.2.1 Psycholinguistic frameworks

Webb (2007b) argued that L2 learners use their L1 knowledge and contextual information when they encounter an unknown lexical item, and that the learning process progresses slowly when the learners lack sufficient information to understand the item. This description appears to be more relevant to L2 learners with limited proficiency, although it is unclear how L2 learners infer the meaning of an item from L1 knowledge or from provided context in the L2.

Pellicer-Sánchez (2015) explained the four measures of processing time most frequently used in eye-tracking studies: (a) first fixation duration, (b) first pass reading time, (c) fixation count, and (d) total reading time. According to Pellicer-Sánchez, first fixation duration and first pass reading time are likely to reflect initial processes in sentence processing that take place at lower levels (i.e., word recognition and lexical access). The number of fixations and total reading times reflect higher order processes, such as semantic integration and comprehension. It appears that learning a novel lexical item involves both low- and high-level processes.

To explain higher-level processing in developing L2 lexical knowledge, Kroll and Stewart (1994) proposed a model for lexical and semantic representations, which they called the revised hierarchical model (RHM). The goal of their experiments was to investigate the structure of bilingual memory, with the assumption that a bilingual mind has two separate lexicons for L1 and L2 at the lexical level, with shared concepts at the conceptual level. According to the model, the L1 lexicon is strongly associated with concepts, because the L1 lexicon develops concurrently with meanings. Although there are assumed links between the L1 lexicon, L2 lexicon, and concepts, the RHM proposes that the strength of these links is likely to change as a bilingual's proficiency changes. The RHM posits a dynamic change in development in the bilingual's mind, describing the importance of the direct link between the L1 lexicon and the L2 lexicon at the initial stage of vocabulary acquisition, and an increase in direct semantic links from the L2 lexicon to concepts as the bilingual's proficiency increases (Kroll & Sunderman, 2003).

By offering a description of the developmental aspect of the bilingual mind, the RHM has become the predominant framework in psycholinguistic research on bilingualism, and the model has been examined in many studies (for reviews, see Brysbaert & Duyck, 2010; Kroll, Van Hell, Tokowicz, & Green, 2010). In particular, the link between the L2 lexicon and concepts has been investigated in conjunction with L2 proficiency. Whether bilinguals with low proficiency show evidence of a conceptual link with their L2 lexicons has been examined using such tasks as translation recognition and semantic categorization (Dufour & Kroll, 1995; Talamas, Kroll, & Dufour, 1999). However, research findings thus far are not conclusive, although they suggest possible links between L2 lexicon and concepts in both lower and higher proficiency bilinguals.

Several interpretations have been offered to understand the mixed findings. For example, Talamas et al. (1999) investigated bilinguals' sensitivity to forms and meanings in a translation recognition task. They hypothesized that less proficient bilinguals would be more affected by forms. However, results showed a form interference effect on bilinguals regardless of proficiency. The interpretation of the results focused on differences in the use of conceptual information, rather than the presence of a conceptual activation link from the L2 lexicon. Talamas et al. suggested that the less proficient bilinguals used any available resources, due to their insufficient L2 knowledge, while the bilinguals at higher proficiency levels used resources available at both the lexical and conceptual levels to perform the task. Therefore, they concluded that conceptual activation patterns are different, depending on proficiency level.

Another interpretation of the mixed results is the possibility of different developmental states at the conceptual level. Dufour and Kroll (1995) hypothesized that concepts contain nodes that could be activated by the L1 alone or by both L1 and L2. In this scheme, the number of nodes that could be activated by both languages increases as a bilingual's proficiency develops. The L2, conversely, begins with a smaller set of nodes. The set is assumed to grow larger as proficiency increases. Therefore, the size of the set was suggested to play a role in conceptual activation in the L2.

Brysbaert and Duyck (2010) questioned "to what extent all semantic information is language-independent, as suggested by the RHM" (p. 367). Consequently, they claimed the need to differentiate language-dependent and language-independent semantic features, noting supportive evidence in the literature, including the memory literature. To this criticism, Kroll et al. (2010) responded that it could be a matter of how common conceptual features are associated with word forms based on the different contexts and structures of language use, rather than the issue of language dependency.

Because the RHM focuses on the developmental aspect of bilingual lexicons and L2 proficiency, the mixed findings of previous studies are difficult to interpret. However, the model provides a distinct level difference between lexical and conceptual representations, which facilitates understanding of the different dimensions of vocabulary knowledge discussed in the IVL literature. In addition, according to the RHM, conceptual activation would be available to low proficiency bilinguals. The model thus supports the idea that L2 learners activate links to the L1 lexicon and concepts when they encounter a novel lexical item in context. The model also implies that incidental learning conditions could lead L2 learners to rely more on conceptual links to search for relevant information in order to understand the meaning of new items.

A caveat concerning the RHM's application to IVL is the model's limited descriptions of lexical level processing. To describe a possible mechanism at the lower level of vocabulary acquisition, including the relationship between L1/L2 words at the lexical entry level and concepts, Jiang (2000) presented a conceptual model for representations of L2 lexical information in the mental lexicon. This framework describes how an L2 learner develops a lexical entry in the L2 lexicon. Jiang considered the process in L2 to be fundamentally different from processes of L1 vocabulary acquisition. Jiang hypothesized three developmental stages: (a) establishing an L2 lexical entry with information about formal specifications, such as phonology and orthography only; (b) amalgamating the information from the first stage and information about the semantics and syntax of its L1 translation; and (c) integrating all the information extracted from exposure to an L2 word as the L2 lexical entry comes to resemble one in the L1. This framework captures an L2 lexical entry's development, although Jiang assumed various stages of development in a learner's L2 lexicon. The second stage of this framework may be closely associated with the RHM, because a novel lexical item begins integrating with available information and L1 knowledge as a representation of the item develops in the L2 lexicon.

Jiang (2002, 2004) examined the second stage of the model, which he calls the L1 lemma mediation hypothesis or semantic transfer hypothesis, using a semantic-relatedness judgment task. The critical stimuli were pairs of English words that share the same translation in English; some had different translations in the participants' L1s while others did not. If participants relied on L1 translation, the pairs with different translations would require longer response times. Results supported the hypothesis, suggesting the L2 participants' reliance on L1 translation. However, with L1 translation, semantic activation must be involved as well. As discussed above, the conceptual links activated by L2 speakers might differ from the conceptual links activated by native English speakers.

The results of both of Jiang (2002, 2004) studies corroborated the RHM by demonstrating links between lexicons and concepts, especially a strong association between L1 words and concepts. The results of other studies (Chen & Leung, 1989; Kroll & Curley, 1988; Kroll & Stewart, 1994; Potter, So, Von Eckardt, & Feldman, 1984) have supported the concept mediation model with higher proficiency bilinguals. However, as the mixed findings suggest, clear explanations of these relationships have not yet been proposed. As Kroll et al. (2010) pointed out, “There is still a debate in the literature as to whether even highly proficient late bilinguals are able to fully access semantic information for the L2” (p. 377).

In short, mechanisms involved in IVL include both higher and lower order processes. When a novel lexical item is presented to L2 learners in context, the learners may search for the L1 equivalent of the item or may use contextual information to infer its meaning. If they have already established links between concepts and L2 words as described in the RHM, the processing speed should be fast. Simultaneously, contextual information may be added to infer the meaning of the lexical item. In other words, when an L2 learner with high proficiency encounters a new word in context through reading and/or listening, s/he may not need to use the links between L1 words and concepts. Direct links between L2 words and concepts may reduce processing time and avoid language switch costs. However, it remains uncertain how proficient an L2 learner must be to establish stable links between L2 words and concepts. The question is whether an L2 learner with low proficiency can do the same without accessing the L1. Even if the learner can process the context in the L2 only, lexical and semantic representations of new words in the context may be fragile and susceptible to decay, due to unstable links between the L2 words and the concepts.

Jiang (2002, 2004) model offers some explanations for the findings reported in the IVL literature. The stages of L2 lexical development in his model begin with a lexical entry that contains a word's formal information, such as spelling and pronunciation. Following the orthography and phonology, morphological information can be added, based on the initial encounters. This may be a reason why IVL studies find greater gains in form recognition. At this stage, the lexical entry lacks a lemma, which consists of syntax and semantics. Information pertaining to the lemma may be available, but is not integrated into the L2 lexicon. Therefore, lexical representation at this stage is weak. The model explains why an L1 translation of the word plays an important role in developing semantics. In IVL, with limited exposures, targeted novel lexical entries might remain at this stage. If L1 translation does not occur in processing L2 input, and if the only available resource is contextual information in the L2, the development of the lexical entry may likewise remain in the L2. Depending on the word, the lemma may be established smoothly, and lexical representation may include conceptual representation. Therefore, differences in responses to productive and receptive form-and-meaning tests might be explained by undeveloped lexical entries.

In this regard, both the RHM and Jiang (2000) models are useful in understanding IVL, as both levels of mental processing should be considered in the development of vocabulary knowledge.

2.2.2 Attention in SLA

To investigate whether IVL has occurred, studies must ensure that a novel lexical item in context attracts an L2 learner's attention. Schmidt (1990, 2001, 2010) proposed the noticing hypothesis: "a hypothesis that input does not become intake for language learning unless it is noticed, that is, consciously registered" (2010, p. 722). Both theoretical and empirical studies in the field of SLA have tested this hypothesis. The findings have not resulted in a consensus due to the complexities of the concept of noticing, and the difficulties of measuring both noticing and intake, among other constructs. Schmidt (2010) clarified the meaning of "consciousness" in the noticing hypothesis by dividing it into intention, attention, and awareness. Unlike incidental learning, intentional learning is goal-directed and deliberate, where L2 learners focus on cues and stimuli for intake. In comparison, incidental learning takes place without particular intention, such as in vocabulary learning through reading activities when the learner's goals are comprehension and enjoyment. Attention refers to "a variety of mechanisms or subsystems, including alertness, orientation, detection within selective attention, facilitation, and inhibition" (p. 724). Schmidt noted that attention appears to be central, but that it is still unclear "whether *all* learning requires attention" (p. 724; italics original).

Tomlin and Villa (1994) emphasized the importance of attention by conducting comprehensive analyses of its subsystems. They focused on three components relevant to SLA: (a) alertness or general readiness to deal with incoming stimuli/data; (b) orientation that facilitates or inhibits detection; and (c) detection that selects, or engages, a particular and specific bit of information. The third component, detection, is what occurs at the key attentional moment for intake of an instance of data, which allows the formulation of a hypothesis about the L2 grammar, and the integration of a lexical item into an L2 lexicon. According to Tomlin and Villa, awareness requires attention, but not vice versa. Their argument is different from the noticing hypothesis in suggesting that learning could occur if a learner's attention is drawn to targets to be learned. The question for instructed SLA is how language instruction can enhance a learner's detection of a piece of specific information, so that it will become intake. One of their suggestions is to provide salient targets, using such techniques as input enhancement. This is the fundamental mechanism that justifies both input enhancement and input modification as useful tools in instruction for SLA.

In conjunction with issues of attention and salience, the effects of input enhancement have been examined in many studies (Issa & Morgan-Short, 2018; Lee & Huang, 2008; Leow & Martin, 2017). Effects on intake have been confirmed, although only at a modest level. Issa and Morgan-Short (2018) employed a picture-sentence matching task, but the results did not show strong evidence of intake. They suggested that input enhancement leads learners to attend to form, but does not lead to deeper form-meaning links.

Given the centrality of learner attention for intake, and the limitations of input enhancement in facilitating deeper processing to establish form-meaning links, an important question is how the quality of the context of learners' encounters with new lexical items can be improved so that greater IVL occurs. In other words, how can instructed SLA provide L2 learners with rich contextual information that facilitates deeper processing and intake of newly encountered vocabulary items as a by-product of primary focused activities?

2.3 Input modification

The literature on input modification consists of two lines of research, one focused on comprehensibility and the other on vocabulary learning, including incidental learning. The notion that input modification facilitates understanding of the modified words or phrases originally comes from studies on "foreigner talk" (FT). At the core of Krashen's (1985) widely known input hypothesis is the critical role of "comprehensible input," which refers to language input slightly more advanced than an L2 learner's level. This is often referred to as the input level of "i + 1," where "i" is the L2 learner's current level and "+1" is the following stage of language acquisition. Long (1980, 1981, 1983a, 1983b) investigated the effects of different types of input, arguing that modifications increase input comprehensibility, which in turn facilitates acquisition. Based on the findings of studies investigating successful cases of language acquisition from comprehensible input, however, what is more important for learning may be "adjustments" that occur during interactions rather than the input itself (Larsen-Freeman & Long, 1991). Such adjustments are made by caretakers talking to toddlers, and by native speakers of a target language talking to non-native speakers whose target language proficiency appears to be lower than that of the native speakers. The goal of the adjustments is interaction: the exchange of meaningful messages and reaching a common ground in understanding during the conversation. Input adjustment, which has also been called input modification, involves both simplification and elaboration (Chaudron, 1983). Elaboration originally referred to modifications of interaction involving features of conversation or discourse functions (Parker & Chaudron, 1987). Chaudron (1983) found elaborated topic reinstatements to affect the recall of topics more strongly than did syntactically simplified input. Because elaboration includes the redundancy and clarification found in FT, research interest shifted to the relationships between input modification, comprehension, and the learning of grammar and vocabulary in conjunction with interaction and auditory input.

Studies on comprehensibility have primarily investigated whether modified input enhances comprehension or receptive skills. Yano, Long, and Ross (1994) summarized previous studies' findings on the effects of input modification: (a) linguistic simplification improves comprehension, although simple sentences alone could interfere with comprehension; (b) simplification does not consistently outperform elaboration; (c) the advantages of elaboration for listening comprehension are consistent; (d) input modification (either simplification or elaboration) is more useful for L2 learners of lower proficiency; and (e) a combination of modification types, such as shorter sentences, repetition, or making topics salient, is more effective than simplification for improving the comprehensibility of passages and lecturettes.

To determine the relative effectiveness of pure simplification and elaboration for reading comprehension, Yano et al. (1994) conducted a large-scale study with 483 EFL university students. They created 13 reading passages of various lengths, each in three versions: genuine (i.e., unmodified), simplified, and elaborated. The simplified version was created by making sentences shorter and minimizing the number of multisyllabic words and embedded clauses in each passage. The elaborated version was created by paraphrasing or providing definitions of low-frequency content words. The three versions (genuine, simplified, and elaborated) differed in readability scores, complexity, and length (number of words in a passage). The complexity of the elaborated version was nearly double that of the original (genuine) version by these three indicators.

Long and Ross (1993) conducted a propositional analysis of the three versions of one of the passages, showing that the number of retained information bits changed, depending on the modification type. A considerable amount of information was lost through simplification, whereas it was retained by elaboration. The elaborated version in fact increased the number of predicates and arguments. The dependent variable in the study was scores on comprehension questions, of which there were three types: (a) replication, which required participants to fill in missing elements that could be found in the passages; (b) synthesis, which required participants to synthesize a number of facts found in the passages; and (c) inference, which required making an inference about the implications of information in the passages, based on the readers' understanding (including background knowledge). The question types were based on a taxonomy developed by Davey (1988). The treatment procedure was conducted in intact classes, but with students randomly assigned to treatment groups by distributing three different types of packets of reading passages (genuine, simplified, elaborated). The participants completed 14 comprehension questions based on the reading passages. The analysis included the participants' scores on an English proficiency test as a covariate. Results showed that the scores of those who read the simplified version were significantly higher than the scores of those who read the baseline (genuine) version, both on overall reading comprehension and in responses to replication items. The elaborated version group scored higher than the baseline group in reading comprehension, but not statistically significantly so. No difference was found in the scores on the synthesis items. On the inference items, however, the scores of those who had read the elaborated version were significantly higher than the scores of the other two groups. This result suggests that the participants who read the elaborated version could successfully link the details of information available in the text to correctly respond to the inference items, even though processing the elaborated version imposed an additional burden due to its redundancy and complexity.

The finding in Yano et al.(1994) that the students who read the elaborated version outperformed the other groups could be considered supportive evidence for an elaboration theory, such as the one Reder (1982) had proposed, which was based on theories of memory and text processing in the fields of cognitive science and educational psychology. In this context, elaboration refers to "any additional facts about material to be remembered (TBR) that are thought about at the time that TBR material is studied," which "can be either generated by the rememberer or presented to the rememberer along with the TBR material" (p. 212). Elaboration theory pertains to "depth of processing" (Craik & Lockhart, 1972, p. 675) and explains that redundancy of information in material helps the rememberer establish a richer knowledge base. Furthermore, Reder argued that a rich knowledge base facilitates smoother learning of subsequent information. Due to the materials used, her experiments did not support the theory; however, she suggested that additional information in an embellishment should be connected to the central message of the information given in a passage. In her conclusion, Reder emphasized that elaboration facilitates deeper processing of incoming information, so that it can be reconstructed and inferred, which Reder described as “elaborative processing” (p. 212) that should facilitate deeper learning.

By adding proficiency as a variable, Oh (2001) expanded the study by Yano et al. (1994) on the effects of the two types of input modification (simplification and elaboration) on reading comprehension. The study design had three modification types (simplification, elaboration, and baseline) and two levels of proficiency (high vs. low). Passages were modified by paying special attention to such factors as linguistic complexity, sentence length, total numbers of T-units, words per sentence, and word frequency. Specifically, the elaborated version was created by adding synonyms and definitions of low-frequency words, repetition of original information, paraphrases, redundancy, and specific examples for clarity of thematic structure. The comprehension test used in the study consisted of three item types: (a) general comprehension (synthesis), (b) specific comprehension (replication), and (c) inferential comprehension. One hundred eighty Korean college students were randomly assigned to six groups, two for each of the three different types of reading passage. Results supported the findings on comprehension in the original study by showing statistically significant differences between the elaboration and baseline groups, regardless of proficiency level. On the inferential items, both high and low proficiency groups with elaboration scored statistically significantly higher than the baseline groups, although there was no such difference between the elaborated group and the simplified group at either proficiency level. On the specific (replication) items, the scores of both proficiency groups showed statistically significant differences from the scores of the baseline groups. On the general (synthesis) items, only the scores of the high proficiency groups showed statistically significant differences from the scores of the baseline groups. It should be noted that the high proficiency group with simplified passages gained the highest total scores on comprehension as well as the highest scores on general and specific items. Nonetheless, the findings suggested that the elaborated passages provided not only readers at the high proficiency level, but those at the lower proficiency level, with more information and, thus, a deeper understanding of the passages.

For learners, simplification may be more accessible, due to the lower cognitive load, while elaboration may make higher cognitive demands. However, it is unclear whether elaboration is more effective for students with larger working memory (WM) capacities. To find more evidence of the possible effects of input modification, investigating the relationship between types of input modification and cognitive capacities, such as WM, in vocabulary acquisition is essential, and will potentially reveal more about the links between comprehension, inference from context, WM, and IVL.

In IVL research, an eye-tracking study conducted by Godfroid et al. (2013) investigated the relationship between attention to TWs during L2 reading and post-test recognition scores by operationalizing attention as a participant’s eye-fixation duration. Study participants were 28 advanced English as a foreign language (EFL) students who read 20 short paragraphs in which 12 paragraphs containing 12 TWs (nine pseudo-words, and three known words as controls) were presented in four conditions: (a) a known word (control condition), (b) a pseudo-word (e.g., PANILINES), (c) a pseudo-word followed by a known word (e.g., PANILINES OR BOUNDARIES), and (d) a known word followed by a pseudo-word (e.g., BOUNDARIES OR PANILINES). The third and fourth conditions, (c) and (d), were hypothesized to provide a strong appositive cue, particularly because they added the explicit marker “or” between the two words, which would help readers directly connect the two words while reading. The eye-tracking data showed longer fixation times on pseudo-words than on known words, and indicated that the appositive cue conditions were the most effective at drawing participants’ attention. Analyzing both critical and post-critical areas of gaze duration, the study suggested that longer gaze duration times at the critical area in the appositive cue conditions reflected sentence-integration processes. In these processes, a mental representation of the novel word was assumed to be updated with details, using knowledge of known words (i.e., the appositive cues). However, a statistical analysis showed no vocabulary gains in any condition. Possible reasons for the results were: (a) that the frequency of the TWs (only one occurrence of each) was too low, and (b) that the timed, unannounced multiple-choice post-test, which had 17 distractors for each TW, was too difficult for the participants.

Godfroid et al. (2013) used a specific type of input modification to increase the salience of TWs. The modification was added adjacent to the TWs, using known words (conditions [c] and [d] above). Although the post-test results did not provide evidence of vocabulary gains, the eye-tracking data revealed that the input modification of the appositive cue (condition [c], i.e., a pseudo-word followed by a known word) induced the longest gaze duration, suggesting deeper semantic processing. This text modification type was adopted from a study by Watanabe (1997). Watanabe compared the effects of marginal glosses, multiple-choice glosses, and appositive cues. Results favored both gloss types over appositive cues and no modification. The marginal gloss group scored the highest, followed by the multiple-choice gloss group. There was no statistically significant difference between the two gloss groups. Watanabe pointed out that participants did not perceive the appositive cues as providing clear explanations of the TWs, possibly because the TWs and the adjacent explanatory words were connected only by commas. Therefore, Godfroid et al. (2013) made the connection clearer by using the coordinating conjunction “or” to signal the explanatory relationship between the words.

Loschky (1994) investigated the effects of three types of auditory input on vocabulary and grammar learning. The underlying research question was whether there was a relationship between comprehensible input and SLA. Loschky used Japanese locational constructions as target structures and investigated whether comprehension was achieved with the help of auditory input. The input types were: (a) baseline unmodified; (b) pre-modified, including both simplified and elaborated; and (c) interactive, which allowed the participants to interact with interlocutors about questions. One measure was an aural vocabulary recognition test, in which the participants selected either “yes” or “no” in response to whether they had encountered stimuli words, including the TWs, during the treatment. The other measure was a sentence verification test, in which the participants selected either “true” or “false” as they judged whether stimuli sentences and accompanying pictures matched. Results confirmed the effects of interaction, and thus supported the interaction hypothesis (Long, 1981, 1983a, 1996). There was no statistically significant difference between the scores of the pre-modified input group and the baseline group, although the pre-modified input group scored lower on both tasks. The poor comprehension of the pre-modified input group appeared to be due to the input not being “comprehensible.” The input that the group received was a mixture of simplified and elaborated input. In addition, the participants, whose Japanese proficiency was at the beginning level, were exposed to the TWs only once. Therefore, even the modified input may have been too difficult for them to comprehend.

Toya (1992) investigated the effects of the explicitness and implicitness of input elaboration in listening. Results of post-treatment vocabulary tests showed that both elaboration conditions led to better vocabulary scores compared to the baseline, but the effects of explicit elaboration were stronger than those of implicit elaboration. Possible reasons for this finding are: (a) the explicit elaboration included repetitions of the TWs, and (b) the instructions to participants may have created an intentional learning condition, as the vocabulary tests were announced prior to the treatment. No effects, however, were found with delayed post-tests.

Chung (1995) investigated the effects of input modification on reading comprehension and incidental vocabulary learning. Twenty TWs were selected with the expectation they would be unknown to the participants, who were English learners in Korea. Frequency of the TWs was not mentioned in the study, and reading materials included both lexical and structural elaboration. The study described "lexical elaboration" as "adding redundancy to the language items anticipated to be unknown to the subjects" (p. 38), and "structural elaboration" as

> adding redundancy to text structure in order to clarify message content and organization through signaling of an inter-sentential relationship, retention of full NPs, repetition, supplying omitted elements, using anaphoric rather than cataphoric reference, and paraphrasing for summary

> statements which make already existing logical relations explicit without adding new information. (pp. 38–39)

Nine genuine passages with various lengths and topics were used as a baseline. Four modified versions (simplified, lexically elaborated, structurally elaborated, and lexically and structurally elaborated) were developed on the basis of the genuine passages. Vocabulary gains were assessed with three post-tests. Immediate tests were a form-recognition test and an L1 meaning-recognition test, with another L1 meaning-recognition test administered one week later to examine longer-term retention. Results showed that the modified version groups all scored higher than the baseline group, although there was no statistically significant difference among them. The group that read the simplified version scored the highest. The mean vocabulary test scores of all the groups fell between 36.1% and 50.3%, with the modified version groups scoring higher on the vocabulary tests than the baseline group. The group that read the simplified version scored highest for both form and meaning. Again, however, these differences were not statistically significant. Possible reasons for the low scores on the post-tests were proficiency level, only a single exposure to each TW, and the format of the post-tests. The tests, which required the participants to choose 20 forms and 20 L1 meanings out of 40 choices, may have been too hard for the participants.

Urano (2000) investigated facilitative effects of input modification on both reading comprehension and IVL. The study used three different types of input modification in reading: (a) baseline, (b) simplified, and (c) elaborated. Stimuli were at the sentence level, and elaboration was operationalized as lexical modification, or adding synonyms to TWs. The dependent variables for comprehension were mean reading time and scores on reading comprehension questions, and for IVL, form- and meaning-recognition scores. Regarding comprehension, the results indicated positive effects of lexical simplification and elaboration on reading times, with no significant effects on scores for the comprehension questions. As for IVL, the results of the form-recognition test and the meaning-recognition test showed that lexical elaboration facilitated IVL more than simplification. In addition, students with higher L2 proficiency benefited more from the elaborated input. The findings of positive effects of lexical elaboration on both form- and meaning-recognition contrast with the findings of a few other IVL studies that have used sentences as context.

Kim (2006) investigated the effects of lexical elaboration and typographical enhancement (i.e., boldfacing) of written input on IVL. Kim's definition of lexical elaboration, "giving learners the meaning of a word" (p. 349), followed those of Urano (2000) and Watanabe (1997). Two hundred and ninety-seven EFL university students in intact classes were randomly assigned to treatment groups in which they received one of six different types of reading passage. Kim administered two vocabulary post-tests, which assessed form-recognition and meaning-recognition. Additionally, he included a retrospective vocabulary pre-test, in which participants reported their knowledge of TWs before the treatment session. This test was administered after the two vocabulary post-tests, and these test scores were used to adjust the post-test scores when necessary. Results of a MANOVA showed no interaction effects between elaboration and enhancement. The findings were as follows: (a) explicit elaboration was significantly more effective for meaning-recognition compared to the baseline passage; (b) explicit elaboration was more effective for meaning-recognition than implicit elaboration; (c) explicit elaboration alone was not effective for form-recognition; and (d) typographical enhancement had significant effects on form-recognition, as predicted, and no effects on meaning-recognition. The findings of an enhancement effect appear to be compatible with findings from previous studies (Lee & Huang, 2008; Leow & Martin, 2017). Kim suggested that the type of typographical enhancement (i.e., boldfacing) might have been responsible for the weak effects on meaning-recognition. The findings on the effects of explicit and implicit elaboration suggest the importance of using clearer signals for lexical elaboration, such as connectors like “which means” before definitions or synonyms. The implicit elaboration of this study was appositive cues following TWs. The effects of the implicit elaboration were found to be weak, which aligns with the findings in Watanabe (1997).

O’Donnell (2009) investigated the effects of elaborative modification on reading passages from Spanish textbooks, compared to the original unmodified version. The purpose of the study was to assess the effects of textual elaboration on reading comprehension and vocabulary recognition using literary texts. O’Donnell did not include a simplified version, claiming that it did not seem possible to maintain the integrity of a literary work that way. Three passages were selected from textbooks to create unmodified and elaborated versions, with text lengths of the elaborated versions increased by an average of 65%. Glossed words from the original text remained in the unmodified version. Dependent variables were the total number of words used in an immediate recall test, along with scores from a vocabulary recognition (L1 translation) test. The recall test results revealed that participants who had read the elaborated version were able to recall significantly more than those who read the unmodified version. The results of the vocabulary test also showed significant effects of the elaborated version. However, O'Donnell noted that overall scores were not high, ranging from 21% to 42%.

Taken together, these studies document beneficial effects of input modification. As in the IVL literature, most of these studies used reading passages. Simplified versions allowed participants better comprehension, due to the use of shorter sentences and familiar lexical items. Yano et al. (1994) and O'Donnell (2009), however, pointed out that simplified versions lose contextual richness and provide less information, as compared to genuine texts and elaborated versions. As the results of studies with inference questions show, participants who read elaborated versions receive additional information. Oh (2001) suggested that readers of elaborated versions might have "more opportunities to process critical information within the text and thus to comprehend the text better, even though the resulting text remains at a high level of linguistic complexity" (p. 86).

Previous studies have centered on two types of elaboration: lexical and structural. Urano (2000) demonstrated the effects of lexical elaboration at the sentence level. However, given the importance of providing richer contextual information, both lexical and structural elaboration could be even more useful than either of them alone. In the IVL and IM research, some methodological issues have been recurrent, particularly (a) treatments providing only one exposure to TWs and (b) participants at low target language proficiency levels (Chung, 1995; Loschky, 1994).

Long (2015, 2019) proposed modified elaboration as a solution for long sentences and utterances in elaborated input. Because input elaboration can be achieved by adding redundancy and regularity to genuine input, elaborated input is longer; that is, it contains more words, sentences, and elements of syntactic complexity, such as S-nodes. Modified elaboration preserves the characteristics of the elaborated version other than sentence length. As Long (2019) explained modified elaborated input, it "exposes learners to nativelike L2 use, increases comprehensibility by retaining the redundancy and other features typical of elaboration, and restores normal sentence length and reasonable syntactic complexity" (p. 12). Whether the effects of such modified elaboration differ from those of modified elaboration, however, has not yet been examined empirically.

Furthermore, previous work on input modification includes investigations of its effects on both comprehension and IVL. However, it remains unclear whether there is a relationship between comprehension and IVL. When the L1 is unavailable, L2 learners rely on context more, and rich contextual information is likely to support understanding of the information attached to new lexical items because L2 learners' attention is more likely to be drawn to such context. If elaboration increases the salience of new lexical items and provides rich contextual information, it may help compensate for limitations of the L2 learning environment (Laufer & Hulstijn, 2001).

2.4 Listening comprehension

Fewer studies have focused on IVL during listening than on IVL during reading. De Vos et al. (2018) conducted a meta-analysis of 32 studies of IVL from spoken input, most including multiple modalities (e.g., reading and listening, or listening with visual cues), but few focused on listening alone. In input modification research, several early studies investigated interaction between native speakers and non-native speakers, using auditory input, such as academic lecturettes. Gradually, however, the number of studies focused on reading has increased.

Vandergrift and Baker (2015, p. 392) adapted three requirements presented by Buck (2001) for defining L2 listening comprehension:

> (a) the ability to process extended samples of realistic spoken language automatically and in real time;
> (b) the ability to understand the linguistic information that is unequivocally included in the text; and,
> (c) the ability to make whatever inferences are unambiguously implicated by the content of the passage.

Subsequently, Vandergrift and Baker (2015) listed five commonalities in L2 comprehension processes for both reading and listening: (a) the need for receptive language processing, including decoding and interpretation; (b) the use of linguistic knowledge and world knowledge for text comprehension; (c) the involvement of top-down and bottom-up processing to apply knowledge sources to the language input during comprehension; (d) the need for cognitive processing to create a mental representation in memory of what has been comprehended; and (e) the involvement of other factors, such as metacognition and motivation. They also described three unique characteristics of L2 listening comprehension that lead to its greater cognitive demands: (a) the difficulty for a listener of taking control due to the transient nature of auditory language; (b) a heavy reliance on working memory (WM) to attend to the online sound stream, segment it for meaning, and store it in a very short period of time; and (c) the need for greater content sensitivity to allow listeners to attend to innuendos and/or important information signaled by prosodic features, such as stress and intonation.

Vandergrift and Baker (2015) investigated some of the cognitive learner variables pertaining to L2 listening comprehension. Participants were 157 seventh-grade students in French immersion classes, with L1 English and intermediate mid-high L2 French proficiency, according to ACTFL guidelines. Participants took seven tests: L1 listening comprehension, L2 listening comprehension, L1 vocabulary knowledge, L2 vocabulary knowledge, auditory discrimination ability, metacognitive awareness of listening, and working memory capacity. Correlational analyses examining the relationships between L2 listening comprehension and the learner variables of interest revealed the strongest association with L2 vocabulary. Results of the correlational analysis for WM and L1 listening ability did not show a statistically significant correlation. Participants were from three different cohorts, and the data suggested a variability among the cohorts. The correlation coefficients improved after the cohorts were combined in the analysis. Subsequently, a path analysis was performed to find a provisional model of how these variables might interact in L2 listening comprehension. The best-fit model began with auditory discrimination, which positively affected WM. WM positively impacted L1 vocabulary, but not metacognition. Both L1 vocabulary and metacognition positively influenced L2 vocabulary, which is a direct precursor to L2 listening comprehension. Concerning WM, Vandergrift and Baker speculated that the use of the participants' L1, English, for administering the tests might have caused the weak relationship with L2 listening comprehension, although L1 use for administering WM tests is recommended by Linck, Osthus, Koeth, and Bunting (2014), and supported by the WM literature.

Therefore, the interpretation of their findings regarding WM and L2 listening comprehension remains unclear.

Bloomfield, Wayland, Rhoades, Blodgett, Linck, and Ross (2010) reviewed the literature on listening comprehension in the L2. The goal of the review was to develop better tests for L2 listening comprehension assessment, so the report was organized by characteristics of the listener, the passage, and the testing conditions. As with Vandergrift and Baker (2015), Bloomfield et al. first reviewed L2 listener factors that might influence L2 listening comprehension, such as WM and L2 proficiency. Given the robust results demonstrating that WM capacities affect L1 comprehension across contexts and populations, L2 listening comprehension is assumed to be related to WM capacities. Although a few studies have reported high correlations between L1 and L2 measurements of WM (Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993), others have suggested that using the L2 for WM measures could be confounded by L2 proficiency, and that the language selection for WM measures (i.e., whether participants use their L1, L2, or no language to perform a WM task) could affect results. Vocabulary size has also been suggested as an important factor for L2 listening comprehension (Martin & Ellis, 2012; Vandergrift & Baker, 2015). An important finding reported by Vafaee (2016) is that less experienced L2 listeners are likely to rely on top-down processing to compensate for their linguistic weaknesses. Vafaee explained the different sources of information that top-down and bottom-up processing draw on to achieve comprehension. While top-down processing relies on general knowledge and familiarity with the context, bottom-up processing relies on information derived from a perceptual source. Both processing systems are assumed to be active during listening.

Therefore, the information sources of both systems should be considered when providing L2 learners with aural input.

Bloomfield et al. (2010) also reviewed text variables, looking at both quantitative and qualitative aspects. They defined quantity as text length, redundancy, and information density; these factors were reviewed first. Length and information density in aural input had a direct, strong impact on cognitive abilities, such as WM capacities. Greater length and density of information made it more difficult for an L2 listener to comprehend the message. However, redundancy could compensate for such difficulties. Findings showed that redundancy, such as repetition and paraphrasing, improves comprehension. These three factors (i.e., text length, information density, and redundancy) were intricately connected, and possibly interacted with one another. Hence, studies using listening passages must address them carefully, both as individual factors and in combination.

In addition, Bloomfield et al. (2010) considered quality, referring to syntactic complexity, concreteness (as determined by whether a passage is concerned with concrete objects or abstract concepts), directness, word frequency, and the cultural specificity of vocabulary and idioms. They reported that results from studies examining syntactically simplified passages were mixed, and that the literature did not address whether increasing syntactic complexity could affect L2 listening comprehension. Inclusion of negatives was found to increase comprehension difficulty, but at a modest level. The concreteness of a passage is determined by whether it is concerned with concrete objects or abstract concepts. For instance, they reported that passages concerning concrete objects seem to be comprehended more easily, based on research findings on reading (due to a lack of aural input studies). Word frequency, in comparison, has been extensively investigated. Overall, the difficulty of passage comprehension increases when infrequent vocabulary is used. Based on L1 research, WM may be involved when a listener attempts to recognize low-frequency words. On the other hand, culturally specific vocabulary and idioms were also found to affect comprehension. Although this factor may be related to topic familiarity, it could be a hidden cause of variance in listening comprehension when study participants are from different backgrounds and cultures.

Other factors pertaining to passage type and organization, as well as auditory features of a passage, could also affect listening comprehension. Although text variables affecting L2 listening comprehension have been underestimated by language instructors, they are closely interconnected and can compensate for or interact with one another. The mixed results in the literature might reflect this interconnectivity. For example, several internet-based Japanese language reading judgment tools have been developed (Lee, 2016). One of the purposes for such tools is to sort official documents based on difficulty level. The text analysis is based on surface features, such as the number of words, the number of subordinate clauses, word type and token, and their frequency. However, due to lack of other information on the texts, such as redundancy, concreteness, information density, and semantics, such tools might not be accurate in predicting comprehensibility according to the IM literature. In any case, the tools are available only for written texts, and further investigation of factors affecting L2 listening comprehension is needed.

2.5 Working memory in L2 listening

As discussed in Section 2.4, L2 listening is more cognitively demanding than L2 reading. Although the literature relating to L2 listening comprehension, phonological short-term memory (PSTM), and working memory (WM) is still limited, it points to significant involvement of both PSTM and WM (Bloomfield et al., 2010; Révész & Brunfaut, 2013; Vafaee, 2016; Vandergrift & Baker, 2015). Recent studies have reported the critical role of memory in L2 vocabulary learning (e.g., Malone, 2018; Martin & Ellis, 2012; Masoura & Gathercole, 1999).

According to Baddeley (2003, 2015), the seminal paper written by Craik and Lockhart (1972) contributed to the understanding of levels of processing and semantic memory, and strongly affected the evolution of the theoretical model proposed by Baddeley and Hitch (1974) by positing that WM has multiple components. Their model proposes that the phonological loop has a direct association with the learning of novel lexical items and is a strong predictor of vocabulary acquisition. Baddeley (2015) explained how the multi-component model has evolved over time by incorporating empirical findings. The original three components were the central executive, the phonological loop, and the visuo-spatial sketchpad. After numerous tests, Baddeley's current model (as of 2015) suggests that the phonological loop assists in learning new words by accessing existing words available in long-term memory. Also, as the name suggests, the phonological loop specializes in temporary storage of acoustic and language stimuli, and the capacity of the phonological loop is referred to as phonological short-term memory (PSTM). In the current model, PSTM is considered one component of WM, although it is occasionally examined as an independent cognitive capacity for language learning, apart from the rest of WM (Gupta & Tisdale, 2009; Linck et al., 2013; Martin & Ellis, 2012). The consensus, however, is that PSTM plays a critical role in new word and language learning. Although both reading and listening require WM capacities, research has suggested that different components may have different functions, resulting in possible differences in association between measurement and ability (Daneman & Carpenter, 1980; Kormos & Sáfár, 2008).

Kormos and Sáfár (2008) investigated different roles of PSTM and WM at different levels of L2 proficiency. They assessed participants' ability in reading, writing, listening, speaking, and use of English. Participants were 121 EFL students between 15 and 16 years of age, enrolled in an approximately 10-month intensive English language program. One cohort comprised beginning students, and the other, students at pre-intermediate levels. English scores were those on a Cambridge First Certificate language exam. For WM measures, a non-word span test was used for PSTM, and a backward digit span test for general WM capacities. Both tests were administered in the participants' L1, Hungarian. Results supported the hypothesis that PSTM has different roles for lower and higher proficiency learners. Statistically significant correlations with L2 writing, use of English, and total proficiency were found in the case of the pre-intermediate group, but not the beginner group. There was no correlation between the scores of the non-word span test and the backward digit span test. Kormos and Sáfár suggested that these results provided evidence of two distinct constructs of PSTM and general WM. With regard to listening, their analysis showed a significant correlation with scores on the backward digit span test, but not with those on the non-word span test.

The seminal study written by Daneman and Carpenter (1980) revealed the limitations of conventional word span WM measures and offered new evidence for a distinct component of WM capacities. They proposed a new measure, called a reading/listening span test, to examine WM capacities for both processing and storage. Their participants were 20 L1-English students. Their scores on the new measures correlated highly with those on both reading and listening comprehension measures. In comparison, the word span measure scores did not correlate with the comprehension scores.

Harrington and Sawyer (1992) included the reading span test along with two other WM measures, a digit span and a word span test, in a study conducted with 34 advanced EFL learners. Results indicated that only the reading span test correlated with grammatical knowledge and reading comprehension measures. Furthermore, there were significant correlations between scores on the WM measures in L1 and L2. Although the format of the test was not complex, it successfully tapped into the capacity for general cognitive processing and language comprehension.

The relationship between WM and vocabulary learning has been reported in both the L1 and L2 literature (Gathercole, Service, Hitch, Adams, & Martin, 1999; Malone, 2018; Masoura & Gathercole, 2005). Martin and Ellis's (2012) investigated WM and PSTM to find evidence of a clear distinction between the two. They defined WM as "both storage and processing of information, measured by reading or listening span tasks" (p. 380) and PSTM as "storage alone, measured by non-word repetition or non-word recognition" (p. 380). To measure PSTM, they used a non-word repetition test and a non-word recognition test; and as a complex WM measure, they used a listening span test

(Daneman & Carpenter, 1980; Harrington & Sawyer, 1992). With 40 native, monolingual English speakers, they conducted artificial language training to investigate the correlation between the above cognitive measures and vocabulary and grammar learning outcomes. Results of their correlation analyses clearly showed no association between PSTM and WM, suggesting that they are separable. Overall, the results showed a significant relationship between PSTM and vocabulary knowledge during the final phase of the study, which aligned with the results in the literature. WM, however, did not correlate with their vocabulary comprehension measure based on accuracy scores on an English translation test. Martin and Ellis found a stronger positive relationship between PSTM and vocabulary learning than between general WM and vocabulary learning. The results also suggested that vocabulary learning is important, even for grammar learning, by validating the links between PSTM and vocabulary learning and between PSTM and grammar learning. Moreover, WM measured by non-word recognition showed a significant relationship with vocabulary learning. The correlation may have been stronger with non-word recognition, a comprehension task, than with non-word repetition, which was a production task, because the input was aural.

Wright (2009) investigated correlations between variations in L2 English development assessed by oral production, grammaticality judgments, and WM. Eleven L1-Chinese learners of English attending study abroad programs in the United Kingdom were assessed on their acquisition of wh-movement over time. For WM capacities, three tests were administered: digits back, story recall (used as a complex WM measure and conducted with L1 aural input), and word and sentence span tests. Although Wright claimed that the WM tests administered in the students' L1 showed the strongest correlations with their scores on a linguistic development test, overall results did not show a significant correlation with other measures, possibly due to the small number of participants.

Linck et al. (2014) conducted a meta-analysis of results from studies that investigated relationships between WM and L2 processing and proficiency outcomes. The data contained 79 independent samples with 748 effect sizes and 3,707 participants. The study's intention was to clarify whether any research design factors might have confounded the results in the previous studies. The specific goals were: (a) to obtain population estimates of the correlation between WM capacities and L2 proficiency; and (b) to investigate relevant variables that might be affecting WM effects, such as WM tasks, L2 outcome measures, and participants' proficiency. All participants in the studies included in this meta-analysis were late adult bilinguals with no history of neurological or psychopathological problems.

The meta-analysis found (a) significant positive relationships between WM and L2 proficiency, processing, comprehension outcomes, and production outcomes; (b) significant relationships between WM and participants at both high and low proficiency levels; (c) larger correlations between WM and L2 proficiency when the L2, not the L1, was used to measure WM; (d) significantly stronger correlations of complex WM span tasks and L2 proficiency than of simple span tasks and L2 proficiency; and (e) significant positive relationships between short-term memory measures, including PSTM. Linck et al. (2014) argued that L2 processing and SLA theoretical frameworks ought to include WM, although they noted that some researchers still disagreed, due to a lack of consistency in the literature. In particular, Linck et al. emphasized a possible larger role of the executive control component of WM than PSTM when using an L2, pointing out that an L2 speaker is required to coordinate multiple cognitive tasks, such as updating incoming information and inhibiting unnecessary language systems. Therefore, they argued, WM tasks that tap into the executive control component of WM may be better predictors of L2 processing and proficiency.

Chapter 3 The Current Study

3.1 Purpose of the study

The aim of this dissertation research was to identify relationships between four kinds of spoken texts—genuine, simplified, elaborated, and modified elaborated—and incidental learning of L2 vocabulary items (L2 IVL).

Many IVL studies have focused on IVL through reading (Godfroid et al., 2013, 2017; Pellicer-Sánchez, 2015; Waring & Takaki, 2003; Webb, 2007a, 2007b, 2008). Mixed modality conditions, such as reading-while-listening, have also been investigated (Brown et al., 2008; De Vos et al., 2018; Hatami, 2017; Malone, 2018; Vidal, 2011), but relatively few studies have investigated IVL through listening alone (Van Zeeland & Schmitt, 2013).

In the input modification literature, it has been over 25 years since listening passages were last used to investigate the effects of input modification for comprehension (Toya, 1992). Most studies have instead used reading passages, and focused on types of written input modification, including simplification and elaboration (Chung, 1995; Kim, 2006; Long & Ross, 1993; O'Donnell, 2009; Oh, 2001; Urano, 2000; Yano et al., 1994). Moreover, few previous studies have explored individual differences, such as WM capacities, in conjunction with IVL and input modification, although the literature has shown that cognitive factors like WM can be influential in language processing.

This study employed four types of spoken input: (a) genuine, (b) simplified, (c) elaborated, and (d) modified elaborated. The last type, modified elaborated, is a new addition to this line of research (Long, 2015, 2019). This study investigated whether the different types of input facilitate listening comprehension. Where comprehension is concerned, the study sought to replicate, with aural input, the study by Yano et al. (1994), which found a relationship between input modification types and comprehensibility through reading. A major goal was to explore whether IVL can be facilitated by providing rich contextual information through input elaboration. The input modification literature suggests that input elaboration supports input comprehensibility through the provision of additional details (e.g., via appositional phrases) concerning lexical items and propositions in a text. It has long been suggested that improved comprehensibility enhances L2 language acquisition (Long, 1980, 1981, 1983a, 1983b).

The IVL literature has investigated the role of context by using a variety of text types, such as graded readers (Nation, 2001, 2013; Webb, 2008). Laufer and Hulstijn (2001) emphasized the cognitive aspect of IVL, suggesting that better retention could be achieved by increased attention to the formal and semantic aspects of words and the rich associations between new lexical items and existing knowledge. Research findings clearly indicate that learning a new word's form occurs more quickly than learning the new word's meaning (Chen & Truscott, 2010; Malone, 2018; Van Zeeland & Schmitt, 2013; Waring & Takaki, 2003; Webb, 2007a). Jiang (2000) argued that the role of L1 translation is critical when learners are establishing a mental representation of a new word with a meaning in the L2 lexicon. Under IVL conditions, the importance of the role of context increases, because a learner uses contextual information to understand the meaning of a new lexical item. Whether L1 translation occurs in the process is unclear.

TWs may require added saliency to draw learners' attention. It was posited that once TWs are detected, IVL can occur; that is, learning can happen without the need for conscious awareness (Schmidt's [2010] "noticing") at the moment the word is encountered. Input enhancement appears to be insufficient to attract learner attention to the semantic level (Issa & Morgan-Short, 2018; Lee & Huang, 2008; Leow & Martin, 2017). Therefore, input elaboration may be more effective for both attracting learner attention and providing rich information associated with new lexical items. This study further explored the roles of WM and L2 proficiency that may account for IVL and comprehensibility in conjunction with individual differences of cognitive processing.

The research questions covered both areas: comprehensibility and IVL. First, comprehensibility, an important requirement in input for language acquisition, was explored in conjunction with the types of auditory passages. Comprehension through hearing the three modified versions was compared with comprehension through hearing the baseline version. The comprehension questions included three types of items: (a) replication, (b) synthesis, and (c) inference. Through responses to these items, the differences in the information that the types of passages provided were explored. Whether L2 proficiency and WM moderate listening comprehension differently depending on passage type was examined separately.

Second, L2 IVL was explored in conjunction with the four types of auditory passages. IVL through listening to the three modified versions was compared with IVL through listening to the baseline version. IVL was operationalized by scores on three types of immediate vocabulary post-test: (a) a form-recognition test, (b) a meaning-recognition in sentences test, and (c) a meaning-recognition with L2 definitions test. Using these measures, the different dimensions of vocabulary knowledge were examined. Additionally, whether L2 proficiency and WM moderate IVL differentially depending on passage type was investigated.

Although part of this study aimed to replicate previous research, most of the research questions were exploratory. The participants were L1-Chinese speakers who were advanced speakers of Japanese. One goal of IVL studies is to investigate how advanced learners can develop vocabulary knowledge efficiently. Research has shown that intentional learning with explicit instruction is more effective for vocabulary learning than incidental learning (for a review, see Hulstijn, 2003). However, explicit instruction and intentional learning are limited by the amount of vocabulary that can be covered, and the amount of vocabulary that advanced learners need to know is large. This is also true for L1 students, who learn new vocabulary through repeated encounters from written and spoken input. In the case of L2 learners, whose cognitive resources are already burdened by having to deal with a non-native language, input containing detectable signals of new lexical items with rich associative information, rather than repetition of the lexical items, could enhance IVL. Therefore, following the literature, the general hypothesis of this study was that input elaboration would facilitate IVL. In particular, modified elaborated input was expected to be more conducive to IVL and comprehensibility, due to the smaller language processing burden created by its shorter, syntactically less complex sentences.

3.2 Research questions (RQs) and hypotheses (H)

To inquire into the comprehensibility of the four types of modified spoken input and their relation to IVL, this study addressed eight specific research questions, presented here along with the hypotheses that guide the research.

3.2.1 Regarding comprehensibility

RQ 1. Does listening to modified versions of spoken input affect L2 listening comprehension, when controlling for L2 proficiency and WM?

H1: Participants who listen to modified versions of spoken input will perform significantly better on listening comprehension than those who listen to a genuine version of the same input.

RQ 2. Do specific types of input modification affect L2 listening comprehension differentially, when controlling for L2 proficiency and WM?

H2: Participants who listen to a simplified version of spoken input will perform significantly better on replication items than those who listen to other versions of the same input.

H3: Participants who listen to a modified elaborated version of spoken input will perform significantly better on synthesis items than those who listen to other versions of the same input.

H4: Participants who listen to a modified elaborated version of spoken input will perform significantly better on inference items than those who listen to other versions of the same input.

RQ 3. Does L2 proficiency moderate listening comprehension with any of the three types of input modification, when controlling for WM?

H5: Participants with higher L2 proficiency who listen to elaborated input will perform significantly better on listening comprehension than those with lower proficiency who listen to elaborated input.

RQ 4. Does WM moderate listening comprehension with any of the three types of input modification, when controlling for L2 proficiency?

H6: Participants with higher WM who listen to elaborated input will perform significantly better on listening comprehension than those with lower WM who listen to elaborated input.

3.2.2 Regarding incidental L2 vocabulary learning (L2 IVL)

RQ 5. Does input modification affect L2 IVL, when controlling for L2 proficiency and WM?

H7: Participants who listen to a modified version of spoken input will perform significantly better on L2 IVL than those who listen to a genuine version of the same input.

RQ 6. Do specific types of input modification affect L2 IVL differentially, when controlling for L2 proficiency and WM?

H8: Participants who listen to a modified elaborated version of spoken input will perform significantly better on a meaning-recognition in sentences L2 vocabulary post-test than those who listen to other versions of the same input.

H9: Participants who listen to a modified elaborated version of spoken input will perform significantly better on a meaning-recognition with L2 definitions vocabulary post-test than those who listen to other versions of the same input.

RQ 7. Does L2 proficiency moderate L2 IVL with any of the three types of input modification, when controlling for WM?

H10: Participants with higher L2 proficiency who listen to elaborated input will perform significantly better on L2 IVL than those with lower proficiency who listen to elaborated input.

RQ 8. Does WM moderate L2 IVL with any of the three types of input modification, when controlling for L2 proficiency?

H11: Participants with higher WM who listen to elaborated input will perform significantly better on L2 IVL than those with lower WM who listen to elaborated input.

Chapter 4 Methodology

Participants listened to four talks as listening passages. The talks were given by four different academic professionals, who spoke on their research. Each talk, or listening passage, had four different versions, including a genuine version. Each talk included eight target lexical items, for a total of 32 target lexical items. All participants listened to all four talks, with a given participant hearing all four in the same input modification version. Each participant was randomly assigned to one of the counterbalanced list of input modification conditions after taking an online proficiency test. While listening, participants were required to respond to comprehension questions that they heard section by section in a passage. The comprehension questions were designed so that the scores would indicate whether the participants focused on the content of the listening passages, which allowed this study not only to replicate the study conducted by Yano, Long, and Ross (1994), but also to investigate the occurrence of IVL during the treatment. If a participant scored below 50% correct (less than 18 out of 36 points) on the comprehension questions, the participant's other data were not included in the analyses. After listening to the four passages, participants were asked to complete the three types of unannounced vocabulary test. The treatment session concluded with a debriefing session, in which participants were asked whether they focused on the content of the listening passages or on the TWs.

4.1 Participants

The participants of this study were L1-Chinese speakers of Japanese as an L2, with Japanese proficiency levels of intermediate-high and above. Based on an a priori power analysis with an effect size of 0.35, the estimated total sample size for the study is approximately 94.[1] Each participant received compensation of 3,500 yen (approximately US$32.00) per session (150 minutes and longer per session).

Recruitment took place in Tokyo. Flyers for recruitment were distributed to members of a large Chinese Christian church in the center of Tokyo, and information sessions were held with groups of Chinese students and professionals. People interested in participating contacted the researcher through email or text messages.[2] The first participants helped recruit others through a social network called "WeChat," used by many Chinese residents in Japan. The sessions were held in quiet meeting rooms at churches, universities, a dormitory for international graduate students, and a preparation school for Chinese students.

To qualify for the study, participants had to (a) have Chinese (Mandarin) as their first language, (b) have lived in Japan for six months or longer, (c) be 18 years old or older, and (d) report Japanese proficiency at the level of intermediate-high or higher. Most of the participants reported attaining the highest proficiency level (i.e., N1) on the Japanese-Language Proficiency Test (JLPT; The Japan Foundation/Japan Educational Exchanges and Services, 2012), and provided a JLPT certificate, although doing so was not required.

A total of 124 L1-Chinese speakers of Japanese participated in the study. However, the comprehension question scores of nine participants did not reach the

---

[1] The power analysis to estimate the number of participants required for this study was conducted using G*Power version 3.1.9.2.

[2] It would have been preferable to prescreen possible participants for language proficiency and other background information, such as length of residence in Japan, but such prescreening proved impractical due to time constraints. Hence, the researcher invited all volunteers who claimed intermediate-high or above Japanese proficiency to take part in the experiment.

threshold score, and four more participants did not meet the condition for minimum length of residence in Japan. In addition, four participants were excluded due to their scores on one of the tests: one scored below the threshold on the Operation Span Task for working memory (WM), two were outliers on the proficiency test, and one was an outlier on the Shapebuilder WM task. Lastly, data from one participant who began studying Japanese with a private tutor at the age of 9 were excluded. As a result, data from 106 participants were used for the study's primary analyses.

These 106 participants consisted of 32 males (30%) and 74 females (70%) with the mean age of 26.2 (range = 18–44, *SD* = 4.50). The mean age of onset of formal Japanese instruction (AFI) was 19.91 (range = 12–27, *SD* = 3.02). Blocked random assignment was used to divide the participants into the four auditory input groups (G: Genuine, S: Simplified, E: Elaborated, ME: Modified Elaborated) after their proficiency level was ascertained through an online test. There were no statistically significant differences between groups in terms of age or AFI, according to results of Scheffe post hoc tests ($F_{\text{age}} = 0.540$, $df = 3, 102$, $p = 0.656 > 0.05$; $F_{\text{AFI}} = 0.336$, $df = 3, 102$, $p = 0.336 > 0.05$) and independent *t*-tests (Age: $t(51) = 1.154$, $p = 0.254 > 0.05$, $t(50) = 0.407$, $p = 0.686 > 0.05$, $t(53) = 0.967$, $p = 0.338 > 0.05$; AFI: $t(51) = 0.755$, $p = 0.454 > 0.05$, $t(50) = 0.761$, $p = 0.450 > 0.05$, $t(53) = 0.895$, $p = 0.375 > 0.05$). As for length of residence in Japan (LOR), 51 participants (48.1%) had lived in Japan more than one year and less than three years ($1 \leq \text{LOR} < 3$); 25 participants (23.6%) had lived in Japan more than three years and less than five years ($3 \leq \text{LOR} < 5$); 23 participants (21.7%) had lived in Japan more than five years ($5 \leq \text{LOR}$); and seven participants (6.6%) had lived in Japan less than one year (6 months $\leq$ LOR < 1). One participant was a fulltime homemaker, and all other participants were either students at educational institutions or professionals who lived and worked in Japan. All participants completed a questionnaire regarding their language background information. Tables 1 and 2 summarize this information regarding age, AFI, LOR, and job status of the participants.

Table 1

Participant Background: Age at Testing and Age of Onset in Formal Instruction

Group	Age at Testing			Age at Onset of Formal Instruction		
N = 106	Mean	Range	*SD*	Mean	Range	*SD*
Genuine *n* = 27	27.00	22–41	5.05	20.41	13–27	3.08
Simplified *n* = 26	25.62	20–34	3.52	19.81	13–27	2.68
Elaborated *n* = 25	26.44	21–44	4.87	19.72	12–25	3.43
Modified Elaborated *n* = 28	25.75	18–37	4.54	19.68	12–26	2.96

Table 2

Participant Background: Length of Residence and Job Status

Group	Length of Residence				Job Status[a]				
N = 106	6 months	1 year	3 years	5 years	a	b	c	d	e
Genuine	1	13	5	8	7	13	1	5	1
n = 27 (%)	(3.7	48.1	18.5	29.6)	(25.9	48.1	3.7	18.5	3.7)
Simplified	2	15	4	5	7	11	1	5	2
n = 26 (%)	(7.7	57.7	15.4	19.2)	(26.9	42.3	3.8	19.2	7.7)
Elaborated	3	9	8	5	7	10	2	4	2
n = 25 (%)	(12.0	36.0	32.0	20.0)	(28.0	40.0	8.0	16.0	8.0)
Modified Elaborated	1	14	8	5	8	12	3	4	1
n = 28 (%)	(3.6	50.0	28.6	17.9)	(28.6	42.9	10.7	14.3	3.6)

[a] Job status: a = professional, b = graduate student, c = undergraduate student, d = student at Japanese language school, e = other (e.g., research student, auditor, vocational school student)

4.2 Materials

4.2.1 Genuine spoken input and target word selection

To create the materials for the four IVL conditions, a search was made for naturally occurring auditory input that could be used as the genuine passages or texts for this study. The intention was to find spoken texts originally created by native speakers for the purpose of actual communication, not for language teaching (Long, 2015, 2019; Widdowson, 1976).[3]

The initial search for input showed that material from news media, such as radio news, was already simplified or edited before becoming available to the public. The syntactic complexity of such texts was found to be rather low; in particular, the texts rarely included subordinate clauses, although they might include specialized vocabulary items depending on the topic.

Therefore, it was necessary to elicit original input for the study. To do so, the researcher invited eight Japanese academic professionals to deliver talks to Japanese high school students on their subjects of expertise. Four of the talks were utilized as listening passages for the study.[4] The topics were: international development, mosquitoes that do not suck blood, lightning, and heritage language education. The reason the target audience was high school students was to induce the professionals to adjust their talk to be easy to follow by listeners outside of their fields. To control the consistency of the genuine input, guidelines were provided to the professionals (Appendix A). The guidelines, which were sent to them electronically, included specific information, such as the occasion (a session showcasing introductions to ongoing studies in various academic fields), the audience (Japanese high school students who would enter university in the near future), and requirements for the talk (4 minutes, approximately 1500 to 2000 characters in the draft, and no PowerPoint slides). The speakers were asked to submit

---

[3] Genuine texts are distinguished from authentic texts, because genuine texts can be authentic or inauthentic depending on their use. For details of this distinction, see Long, 2015, 2019.

[4] The four talks were selected on the basis of content, vocabulary, and length.

drafts of their talks to the researcher first.[5] The drafts were expected to include relevant data, specific examples, and explanations of any specialized vocabulary, so that the audience would understand the content. If any content or sentences were unclear, the professional was asked to revise the draft. Once the final drafts were accepted, the professionals were notified of the real reason for the drafts. The titles of the talks and the fields of study are presented in Table 3.

Table 3

Titles of Talks and Fields of Study

Talk no. Titles	Fields of study
1. 国際開発学の大切さと将来の仕事 Significance of international development studies and future careers	International Affairs
2. 血を吸わない蚊の生態の研究 Research on the ecology of mosquitoes that do not suck blood	Public Health
3. 雷はどうやって起きるのか How does lightning occur?	Astrophysics
4. ワシントン日本語継承センターの継承語教育について Heritage language education at the Washington Japanese Heritage Center	Language Education

Selection of target words

From each draft of a talk, eight nouns were selected as the TWs of the study. The selection procedure comprised three steps: (a) nouns that were not important to understand the text (non-key-words) were identified, (b) the nouns were checked on a frequency list of Japanese words (Matsushita, 2011), and (c) eight low-frequency nouns

[5] This study counted on the fact that it is customary for Japanese speakers to write drafts prior to delivering speeches.

(at the 10,000-word family level of Nation's [2001, 2013] classification) were chosen. Appendix B provides the list of TWs for the study.

In Nation's (2001, 2013) frequency classification for vocabulary, the 2,000-word families are considered high-frequency, 7,000-word families are mid-frequency, and 10,000-word families and above are low-frequency. This classification is based on the British National Corpus (BNC), which contains approximately one hundred million words. The frequency list used in this study was the Vocabulary Database for Reading Japanese (VDRJ, version 1.1; Matsushita, 2011). The VDRJ contains 32,819,412 words, and was developed based on books, a specific Q&A website on the Internet, and the Balanced Corpus of Contemporary Written Japanese (BCCWJ) created in 2009 by the National Institute for Japanese Language and Linguistics (NINJAL).[6] Although the size of the Japanese frequency list is approximately one third that of the BNC, the study adopted Nation's classification system.

The TW list included some English loanwords. The number of English loanwords in Japanese has been increasing (Matsumoto, 1989/1994). Research has shown, however, that encoding these words in Japanese may be more difficult for non-native Japanese speakers, including L1-Chinese speakers of Japanese (Daulton, 2008). As shown in Appendix B, a few of the English loanwords in the TW list were not found in the VDRJ.

Some of the TWs, such as *zesei* 'correction' in the first talk and *seesoku* 'inhabitation' in the second talk, can be used as verbs when *suru* 'to do' is attached to them (Shibatani, 1990). To avoid any confusion by using different parts of speech, all TWs were used as nouns in all passages. The length of the TWs varied because the

---

[6] NINJAL also developed a spoken corpus. Its availability, however, is limited, according to their website (https://pj.ninjal.ac.jp/corpus_center/csj/).

selection was limited by the original content and vocabulary items used by the professionals. Due to the length of the passages, ecological validity was prioritized for TW selection.

Additionally, the cognate status of the TWs for Chinese speakers was examined during pilot sessions with two L1-Chinese speakers who lived and worked in Japan. The low-frequency words in Japanese chosen for the study were confirmed to be low-frequency words in Chinese, and most of them did not have Chinese cognates; thus, the selection fit the criteria for TWs described by Vidal (2011).

Each TW was embedded three times at natural points in the passages. The embedding locations were determined according to the flow of the talks to maintain ecological validity. The frequency of exposure to the TW was determined by choosing the smallest possible number for potential frequency effects, based on the IVL literature (Van Zeeland & Schmitt, 2013; Webb, 2007a). Due to the low frequency of the TWs, participants would likely hear them in the listening passages for the first time. Their attention was drawn to the content of the passages by the comprehension questions. The three exposures alone were not expected to draw their attention or to override the effects of the input modification. In other words, three exposures were considered sufficient for participants to learn the novel lexical items incidentally. The unannounced vocabulary tests after the listening passages assessed their incidental vocabulary learning of the TWs.

In summary, the genuine texts were created based on four original talks. Each talk had eight TWs and each TW appeared three times. The genuine talks, which had no explanations or definitions of the TWs, were the baseline for all of the modified versions. (See Appendix C for a sample of a genuine version.)

4.2.2 Modified spoken input

This section describes the input materials, while Section 4.2.3 explains in greater detail how they were developed. The three types of modified auditory input (simplified, elaborated, and modified elaborated) were developed on the basis of the genuine listening passages by controlling redundancy and regularity of information. Vocabulary items and syntactic complexity were controlled when necessary. In particular, the number of sentence-nodes (S-nodes) and the number of letters in the texts were controlled to create the simplified and elaborated versions. In prior research, S-nodes typically have been used as an index of speech complexity; they are argued to reflect the syntactic complexity of language by indicating the number of verbs embedded in a sentence (Kobayashi & Rinnert, 1992). S-nodes in this study included verbs and copula.[7]

The spoken input was recorded by male and female native Japanese speakers. Recording took place in a sound-proof recording booth, using high quality recording equipment. Care was taken to create natural-sounding speech at an appropriate speed. The speed of the audio was designed to be 300 letters per minute, in accordance with standards set by the Japanese public broadcasting company, NHK, *Nippon Hoso Kyokai*, or Japan Broadcasting Corporation. This speed is considered to be easiest and clearest for an audience to listen to. Depending on the type of modification, the lengths of the audio files were expected to vary slightly. All audio files were edited using the digital editing program Audacity.

Descriptive statistics of all texts and the means for all versions are shown in Table 4. Total letters, total sentences, the number of letters per sentence, total S-nodes, and the

[7] Tamaru and Yoshioka (1994) reviewed problems with S-nodes in Japanese. However, for the purposes of this study, the S-node as currently defined was considered an appropriate indicator of syntactic complexity.

number of S-nodes per sentence were examined for changes in passages, sentences, and syntactic complexity. In previous studies, the number of words has been used to determine the length of a text. In Japanese, however, the number of letters is used instead. The results showed that the number of letters was comparable to words in English.

Total letters and total S-nodes increased gradually across text types in the order of simplified, elaborated, and modified elaborated. The three other factors, total sentences, letters per sentence, and S-nodes per sentence varied in accordance with the length of sentences in the texts. The number of S-nodes per sentence indicated the syntactic complexity; the highest numbers of S-nodes were in the elaborated versions of all four passages. Although the modified elaborated versions preserved most of the elaborated passages, the numbers of S-nodes per sentence were comparable to those in the simplified versions. This confirmed that the modified elaborated versions effectively segmented the long sentences of the elaborated versions, as intended.

To further investigate differences between the text types, a Mann-Whitney U test was conducted on the means of total letters and the means of total S-nodes of the three modified versions in comparison to the genuine version for all four passages. As seen in Table 5, the results show no statistically significant difference between genuine and elaborated versions. In contrast, simplified versions and modified elaborated versions were significantly different from genuine versions, with a range of effect sizes from small to medium (d = 0.15, 0.48). When all talks were amalgamated, the elaborated versions also indicated a statistically longer length than the genuine versions, with the smallest effect size (d = 0.15).

Table 4

Spoken Input: Descriptive Statistics

		Total letters	Total sentences	Letters per sentence	Total S-nodes	S-nodes per sentence
Talk 1	Genuine	1805	32	56.4	104	3.3
	Simplified	2133	59	36.2	137	2.3
	Elaborated	2344	33	71.0	136	4.1
	Modified Elaborated	2491	61	40.8	145	2.4
Talk 2	Genuine	1657	34	48.7	124	3.6
	Simplified	1836	65	28.2	135	2.1
	Elaborated	2329	39	59.7	172	4.4
	Modified Elaborated	2472	74	33.4	177	2.4
Talk 3	Genuine	1879	37	50.8	153	4.1
	Simplified	2054	70	29.3	164	2.3
	Elaborated	2200	38	57.9	179	4.7
	Modified Elaborated	2338	70	33.4	180	2.6
Talk 4	Genuine	1700	28	60.7	103	3.7
	Simplified	1874	56	33.5	118	2.1
	Elaborated	2173	32	67.9	140	4.4
	Modified Elaborated	2308	60	38.5	145	2.4
Means	Genuine	1760	33	54.2	121	3.7
	Simplified	1974	63	31.8	139	2.2
	Elaborated	2262	36	64.1	157	4.4
	Modified Elaborated	2402	66	36.5	162	2.4

Table 5

Spoken Input: Mann-Whitney U Test Results

		Total letters *w*-statistic	*p*-value	Effect size	Total S-nodes *w*-statistic	*p*-value	Effect size
Talk 1	Simplified	1392.0	0.00***	0.39	1207.5	0.02*	0.24
	Elaborated	428.5	0.19		424.0	0.17	
	Modified Elaborated	1341.0	0.00***	0.31	1227.5	0.04*	0.22
Talk 2	Simplified	1713.5	0.00***	0.45	1576.5	0.00***	0.36
	Elaborated	491.0	0.06		704.5	0.65	
	Modified Elaborated	1755.0	0.00***	0.32	1681.5	0.00***	0.28
Talk 3	Simplified	2046.5	0.00***	0.48	1988.0	0.00***	0.45
	Elaborated	592.0	0.24		580.0	0.19	
	Modified Elaborated	1934.0	0.00***	0.40	1896.5	0.00***	0.39
Talk 4	Simplified	1227.5	0.00***	0.46	1163.0	0.00***	0.40
	Elaborated	395.5	0.44		378.0	0.30	
	Modified Elaborated	1222.5	0.00***	0.36	1172.5	0.00***	0.33
All talks	Simplified	25302.0	0.00***	0.45	23559.0	0.00***	0.37
	Elaborated	7728.5	0.02*	0.15	8387.0	0.16	
	Modified Elaborated	24866.0	0.00***	0.35	23850.0	0.00***	0.31

4.2.3 Developing the modified spoken input

Following study designs described in the IVL literature (Chung, 1995; Long, 2015), three modified versions of the input were created for this study. Due to the length, the content, and the purpose as explained to the Japanese academic professionals who provided the genuine input, the modification affected the texts at a lexical level and at a sentence level.

4.2.3.1 Simplified version

Based on the characteristics of linguistic simplification in pedagogic materials that described by Long (2015), the following guidelines were developed to guide the creation of the simplified passages in this study:

1) Long sentences were divided into two or more shorter ones. As shown in the examples below, each version contains the same information, but the simplified version consists of two sentences. The number of letters in each version is approximately the same (G-56, S-20 + 35 = 55). However, the overall number of S-nodes in the whole simplified text was increased, due to the additional sentence (G-6, S-4 + 3 = 7).

Ex.1-G) 血を吸う昆虫といっても色々あるわけですが、シラミ、ノミ、蚊などは代表的なもので、哺乳類の血を吸って生きています。

> *chi o suu konchuu to ittemo iroiro aru wake desu ga, shirami, nomi, ka nado wa daihyooteki na mono de, honyuurui no chi o suttee ikite imasu.*
>
> Speaking of insects that suck blood, they vary, but (those) such as lice, fleas, and mosquitos are typical, and they live by sucking the blood of mammals.

Ex.1-S) 血を吸う昆虫といっても色々あるわけです。そのうち、シラミ、ノミ、蚊などは代表的なもので、哺乳類の血を吸って生きています。

> *chi o suu konchuu to ittemo iroiro aru wake desu. sonouchi shirami, nomi, ka nado wa daihyootekina mono de, honyuurui no chi o suttee ikite imasu.*
>
> Speaking of insects that suck blood, they vary. Among them, (those) such as lice, fleas, and mosquitos are typical, and they live by sucking the blood of mammals.

2) When necessary, an inter-sentential connector was added at the beginning of the second sentence. The genuine versions contained lists of propositions, often connected using *ga* ‘but’. In such cases, either *shikashi* ‘however’ or *demo* ‘but’ was placed in the initial position in the second sentence.

Ex. 2-G) 長い間、血を吸う昆虫の研究をしてきましたが、今日は、血を吸わない蚊の研究についてお話しします。

> *nagai aida, chi o suu konchuu no kenkyuu o shite kimashita ga, kyoo wa chi o suwanai ka no kenkyuu nitsuite ohanashi shimasu.*

For a long time, I have been studying insects that suck blood, but today, I will talk about research on mosquitoes that do not suck blood.

Ex.2-S) 長い間、血を吸う昆虫の研究をしてきました。しかし、今日は、血を吸わない蚊の研究についてお話しします。

nagai aida, chi o suu konchuu no kenkyuu o shite kimashita.
shikashi, kyoo wa, chi o suwanai ka no kenkyuu nitsuite ohanashi shimasu.

For a long time, I have been studying insects that suck blood.
However, today, I will talk about research on mosquitoes that do not suck blood.

As a result of this step, the number of letters has slightly increased in the simplified version (G-47, S-21 + 29 = 50). The number of S-nodes has remained the same (G-4, S-2 + 2 = 4).

3) When necessary, a pronoun that indicates a preceding word and/or phrase (i.e., *sonna* 'such', *souiu* 'such') was added to make the connection between nouns clear.

Ex. 3-G) 富山ではいつも午前中に観察し、金沢では昨年以前は午後に観察していたという単純な違いが結果を左右していました。

toyama dewa itsumo gozenchuu ni kansatsu shi, kanazawa dewa sakunen izen wa gogo ni kansatsu shite ita to iu tanjunna chigai ga kekka o sayuu shite imashita.

The simple difference (of time) that I always observed in the morning in Toyama, while I observed in the afternoon in Kanazawa before last year, affected the results.

Ex. 3-S) 富山ではいつも午前中に観察していました。しかし、金沢では昨年以前は午後に観察していました。そんな単純な違いが結果を左右していました。

toyama dewa itsumo gozenchuu ni kansatsu shite imashita.
shikashi, kanazawa dewa sakunen izen wa gogo ni kansatsu shite imashita.
sonna tanjinna chigai ga kekka o sayuu shite imashita.

I always observed in the morning in Toyama.
But in Kanazawa, I observed in the afternoon before last year.

Such a simple difference affected the results.

The relative clause of the subject of the sentence, ‘a simple difference’, in the genuine version was lengthy, but it describes an order of events that occurred chronologically. The order was changed in the simplified version. After the events that occurred are stated in simple sentences, ‘a simple difference’ is used as the subject of the third sentence. The pronoun *sonna* ‘such’ is added to it. In this case, the number of letters increased in the simplified version (G-57, S-20 + 27 + 21 = 68), although the number of S-nodes decreased (G-4, S-3). The decrease of S-nodes occurred due to the segmenting of the relative clauses in Japanese. In the genuine version, ‘a simple difference’ is explained with two factual events, and the relative clause in Japanese contains a verb, *iu* ‘say’, as a connector to the noun. This format of the relative clause is natural and frequent in speech by native speakers of Japanese. In the simplified version, it is harder to include such relative clauses, due to the need to avoid complexity. The examples validate that the number of S-nodes serves as an indicator of syntactic complexity in Japanese as well.

4) In the simplified version, some lower frequency words were replaced with higher frequency words, which had a better chance of being recognized and understood.

In Example 4, the genuine version contains four specialized lexical items: *shoo seetai kee* ‘a small ecological system’, *yoochuu* ‘larvae’, *hoshoku-sha* ‘predator’, and *boofura* ‘wigglers’. The last word, *boofura* ‘wigglers’, is a TW for this passage, and therefore remained across versions. Of the other three, two are replaced with higher frequency words as shown below, and one is removed entirely.

When simply exchangeable, a lower frequency word that could be unknown to listeners was replaced by a synonym of higher frequency (see Table 6 below). When a kanji compound word was likely to be unknown, it was replaced by paraphrasing it with several high-frequency words. In the examples below, the specialized words are kanji compound words. When suffixes appeared with a content word, the frequency of the word was examined separately. In (4), the substitute for the word in the genuine version is a combination of one adjective and two nouns. This is a result of unpacking the dense kanji compound word. This combination consists of high-frequency words, as shown. The third specialized word, *hoshoku-sha* ‘predator’, is dropped in the simplified version; the second sentence has already covered the meaning of ‘predator’, when describing how a mosquito larva eats larvae of other kinds. This example illustrates how information in the genuine versions is diluted in the simplified versions.

5) Sentences including overly complex, grammatical structures in the genuine version were simplified by dividing them into shorter sentences for the simplified version. As shown in the same example below, while the number of letters has increased by 11 in the simplified version (G-63, S-27 + 28 + 19 = 74), the number of S-nodes decreased by one (G-6, S-2 + 2 + 1 = 5).

Ex. 4-G) ボウフラは水溜りという小生態系の中にいて、最近の研究でわかったのは、他種の幼虫を餌として食べる捕食者の蚊の幼虫がいることです。

boofura wa mizutamari to iu shoosaitaikei no naka ni ite, saikin no kenkyuu de wakatta no wa, tashu no yoochuu o esa toshite taberu hoshokusha no ka no yoochuu ga iru koto desu.

Wigglers (TW) live in a small ecological system called a water puddle, and what Was found in recent research is that there are larvae of mosquitoes, as predators, that eat other kinds of larvae as food.

Ex. 4-S) ボウフラは水溜りという小さな生き物の世界の中にいます。そこで、他の種類の子供を餌として食べる蚊の子供もいます。このことが、最近の研究でわかりました。

boofura wa mizutamari to iu chiisana ikimono no sekai no naka ni imasu.
sokode hoka no shurui no kodomo o esa toshite taberu ka no kodomo mo imasu.
konokoto ga saikin no kenkyuu de wakarimashita.

Wigglers (TW) live in a world of small creatures called a water puddle.
In the puddle, there are mosquito children that eat children of other kinds of insects as food. This was found in recent research.

Table 6

Lexical Items Changed in Ex. 4-S

Genuine	Simplified
小生態系 (a small ecological system) *shoo seitai kee* 2K 6K(7K) 2K	小さな生き物の世界 (a world of small creatures) *chiisana ikimono no sekai* 1K 4K 1K 1K
幼虫 (larvae) *yoochuu* 20K	子供 (children) *kodomo* 1K

6) English loanwords were replaced with Japanese and kanji-based words. As mentioned above, the number of English loanwords in the Japanese language has been increasing. However, they are harder for non-native speakers of Japanese to understand due to the way they are adapted to Japanese phonology. In addition, the original meaning of an English loanword is sometimes distorted or narrowed in Japanese. Therefore, relatively new English loanwords to Japanese were replaced with Japanese words and/or kanji compound words, as shown below.

Ex. 5-G）一つは工場で働く女性向けのリーダーシップトレーニングの教科書で、

hitsotsu wa koojoo de hataraku josee muki no riidaashippu toreeningu no kyookasho de,

The first one was a textbook for a leadership training course for women working in factories,

Ex. 5-S) 一つは工場で働く女性向けのリーダーを育てるための研修の教科書で、

hitsotsu wa koojoo de hataraku josei muke no riidaa to sodateru tame no kenshuu no kyookasho de,

The first one was a textbook for training women working in factories on how to raise leaders,

7) The topic marker *wa* was added to shorter sentences when appropriate. This was done to improve natural grammatical flow. The information remains the same.

Ex. 6-G) 私の大学は、アメリカのワシントン DC にあって、大学の周りに国際開発のための、国際機関やシンクタンクが結構あります。

watashi no daigaku wa, amerika no washinton dc ni atte, daigaku no mawari ni kokusai kaihatsu no tame no, kokusai kikan ya shinkutanku ga kekkoo arimasu.

My university is in Washington D.C., and there are fairly many international organizations and think-tanks for international development around the university.

Ex. 6-S) 私の大学は、アメリカのワシントン DC にあります。大学の周りには、国際開発のための、国際機関やシンクタンクが結構あります。

watashi no daigaku wa, amerika no washinton dc ni arimasu.daigaku no mawari niwa, kokusai kaihatsu no tame no, kokusai kikoo ya shinku tanku ga kekkoo arimasu.

My university is in Washington D.C. Around the university, there are fairly many international organizations and think-tanks for international development.

4.2.3.2 Elaborated version

The elaborated version was created through changes at the lexical and discourse levels based on findings in the literature (Larsen-Freeman & Long, 1991; Long, 1983a, 1983b, 2015, 2019). In this study, “lexical elaboration” meant adding to a lexical item its definition, synonyms (lexical switches), and/or exemplification to clarify the meaning of the item (Chung, 1995). “Structural elaboration” referred to adding redundancy and regularity to text structure in order to clarify message content and organization by signaling inter-sentential relationships. Redundancy included exact repetitions and repetitions through paraphrasing and/or altering voice. Regularity included retention of canonical word order and dropped pronouns, such as noun phrases. Paraphrasing in summary statements was also used in elaboration. All of these additions presented information in a more logical and explicit way without adding new information (Chung, 1995).

1) Lexical elaboration was added for words that could possibly be unknown to the listeners. As mentioned above, all of the TWs were low-frequency words. Therefore, lexical elaboration was added to them through apposition and explicit signaling, as shown below (Chung, 1995). For other lexical items, non-appositional elaboration was used (see Ex. 8 below).

As a result, the number of letters increased by 50% in the elaborated version (G-66, E-99). The number of S-nodes increased by two in the elaborated version (G-6, E-8).

Ex.7-G) カの雄が花の蜜を吸うことや植物の傷から樹液を吸う断片的な報告がありますが、どの蚊がどの花に集まるのか野外で調べることは難しいのです。

ka no osu ga hana no mitsu o suu koto ya shokubutsu no kizu kara jueki o suu danpentekina hookoku ga arimasu ga, dono ka ga dono hana ni atsumarunoka yagaide shiraberu koto wa muzukashii no desu.

There are fragmentary reports about male mosquitoes’ eating flower nectar and tree sap (TW) by sucking, but it is difficult to investigate which mosquitoes gather at which flowers in the field.

Ex.7-E) カの雄が花の蜜を吸うことや植物の傷から<u>樹液</u>、つまり、木の幹から出てくる水のような液体を吸う断片的な、いくつかの報告がありますが、どの蚊がどの花に集まるのか野外で調べることは難しいのです。

> *ka no osu ga hana no mitsu o suu koto ya shokubutsu no kizu kara <u>jueki</u>, tsumari, ki no miki kara dete kuru mizu no yoona ekitai o suu danpentekina, sukoshi no bubuntekina hookoku ga arimasu ga, dono ka ga dono hana ni atsumarunoka yagaide shiraberu koto wa muzukashii no desu.*
>
> There are fragmentary, or limited, reports about male mosquitoes' eating flower nectar and <u>tree</u> <u>sap</u> (TW), (or) liquid-like-water extracted from tree trunks, but it is difficult to investigate which mosquitoes gather at which flowers in the field.

The following example shows the use of exemplification. Specialized lexical items for classes of animals were used in the genuine version. Specific names of animals were added to exemplify the classes of animals. Non-appositional locations were used in such cases. The number of letters increased in the elaborated version, while the number of S-nodes remained the same (Letters: G-57, E-77; S-nodes: G-2, E-2).

Ex.8-G) 日本には 130 種ほどの蚊の棲息がわかっていますが、ほとんどは吸血性で、ほ乳類、鳥類、爬虫類、両生類の血を吸います。

> *nihon niwa 130shu hodo no ka no seesoku ga wakatte imasu ga, hotondo wa kyuuketsusee de, honyuurui, choorui, hachuurui, ryooseerui no chi wo suimasu.*
>
> In Japan, it is found the living (TW) of approximately 130 kinds of mosquitoes, but most of them are hematophagous, and they suck the blood of mammals (TW), fowls, reptiles, and amphibians.

Ex.8-E) 日本には 130 種ほどの蚊の棲息がわかっていますが、ほとんどは吸血性で、ほ乳類、トリなどの鳥類、ヘビやトカゲなどの爬虫類、カエルなどの両生類の血を吸います。

> *nihon niwa 130shu hodo no ka no seesoku ga wakatte imasu ga, hotondo wa kyuuketsusee de, honyuurui, tori nado no choorui, hebi ya tokage nado no hachuurui, kaeru nado no ryooseerui no chi wo suimasu.*
>
> In Japan, it is found the living (TW) of approximately 130 kinds of mosquitoes, but most of them are hematophagous, and they suck the blood of mammals (TW), fowls, such as chickens, reptiles, such as snakes and lizards, and amphibians, such as frogs.

2) Where information was expressed with overly specialized vocabulary items and syntactically complex sentences, paraphrased information was added, as shown below (Chung, 1995).

Example 9 includes specific names of tropical diseases, which are changed to fit Japanese phonology and followed by a Japanese word, *netsu* ‘fever’. These names may confuse listeners who are not familiar with them. Even for those who may be familiar with the disease names, the change to Japanese phonology could require additional processing time. Moreover, the word for ‘secondary infection’ is a kanji compound word with which listeners may be unfamiliar. The example sentence in the genuine version includes substantial information to process. Therefore, in the elaborated version, a paraphrased summary was added to enhance the presentation of the information. As a result, both the number of letters and the number of S-nodes were more than doubled (Letters: G-69, E-69 + 99 = 168; S-nodes: G-5, E-5 + 6 = 11).

Example 3 is repeated below with the changes made for the elaborated version. The expressions used in the genuine version did not provide a straightforward description of what happened in the course of the observation of the male mosquitos. In the elaborated version, a paraphrasing summary statement was added with more specific and useful information embedded in the genuine version to make it more straightforward. This procedure resulted in the doubling of the letters and S-nodes (Letters: G-57, E-57 + 52 = 109; S-nodes: G-4, E-3 + 6 = 9).

Ex.9-G) 最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告され、国内での二次感染を防ぐための対策を考える必要がでてきました。

Saikin, nettai no byooki to kangae rarete ita dengu netsu, jika netsu nado no *kanjia ga nihon demo hookoku sare, kokunai deno nijikansen o fusegu tame no taisaku o kangaeru hitsuyoo ga dete kimashita.*

Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well, and strategies that prevent domestic secondary infection are becoming necessary.

Ex.9-E) 最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告され、国内での二次感染を防ぐための対策を考える必要がでてきました。わかりやすく言うと、デング熱、ジカ熱などは、熱帯の病気と考えられていましたが、最近これらの病気の患者が日本でも報告されているので、国内でそのような患者から他の人に移らないようにする必要が出て来ました。

saikin, nettai no byooki to kangae rarete ita dengu netsu, jika netsu nado no kanjia ga nihon demo hookoku sare, kokunai deno nijikansen o fusegu tame no taisaku o kangaeru hitsuyoo ga dete kimashita.
wakariyasuku iuto, dengu netsu, jika netsu nado wa, nettai no byooki to kangaerarete imashita ga, saikin korera no byooki no kanja ga nihon demo hookoku sarete iru node, kokunai de, sonoyoona kanja kara hoka no hito ni utsuranai yooni suru hitsuyoo ga dete kimashita.

Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well, and strategies that prevent domestic secondary infection are becoming necessary.
Simply put, it has been thought that diseases, such as dengue fever and zika fever, are those of tropical areas, but recently, patients with these diseases have been reported in Japan; therefore, in Japan it is becoming necessary to make sure these diseases do not get passed from these patients to others.

Ex.3-E) 富山ではいつも午前中に観察し、金沢では昨年以前は午後に観察していたという単純な違いが結果を左右していました。わかりやすく言うと、金沢では蚊が出て来る午前に観察していなかったから、見つけられなかったということです。

toyama dewa itsumo gozenchuu ni kansatsushi, kanazawa dewa sakunen izen wa gogo ni kansatsu shiteita to iu tanjunna chigai ga kekka o sayuu shite imashita. wakariyasuku iuto, kanazawa dewa ka ga dete kuru gozen ni kansatsu shite inakatta kara, mitsuke rarenakatta to iu koto desu.

The simple difference (of time) that I always observed in the morning in Toyama, while I observed in the afternoon in Kanazawa before last year, affected the results. Simply put, the reason I did not find them in Kanazawa was because I was not observing in the mornings when the mosquitoes appear in Kanazawa.

3) When information in the genuine version included elliptical expressions, the missing elements were supplied to clarify the content of the information. Japanese is well-known for a high degree of ellipsis, including “pro-drop” (Shibatani, 1990). The proportion of ellipsis increases in spoken Japanese. In particular, grammatical particles, such as *ga* (subject marker) and *o* (object marker), often disappear. Therefore, in the elaborated version, missing elements were recovered, an example of redundancy, to make more specific information available to the listeners.

Ex.10-G)ウミガメ保護
umigame hogo
sea turtle protection

Ex.10-E)ウミガメを保護する活動
umigame o hogo suru katsudoo
an activity to protect sea turtles

Ex.11-G) ビジネスは彼らの利益になるものの、
bijinesu wa karera no rieki ni naru monono
although our business brought them profits,

Ex.11-E)
私が勤めた日本の会社との貿易ビジネスは貿易取引の相手の利益になるものの
watashi ga tsutometa nihon no kaisha tono booeki torihiki no aite no rieki ni naru monono

although the trading business with the company that I worked for brought profit to our counterparts in a business deal

4) In the elaborated version, pronouns were replaced with specific nouns, as shown in Example 12. Alternatively, as shown in Example 13, specific information (‘for gamma ray measurement’) was inserted between the pronoun and the noun (*gijutsu* ‘technique’). Both types of change added clarity of information. In Example 13, the number of letters

has increased slightly, while the number of S-nodes remains the same (Letters: G-34, E-42; S-nodes: G-4, E-4).

Ex. 12-G) 彼らの支障を取り除く
karera no shishoo o torinozoku
removing their obstacles

Ex.12-E) 途上国の人たちの支障を取り除く
tojookoku no hito tachi no shishoo o torinozoku
removing the obstacles of people in developing countries

Ex. 13-G) 私はこの技術を応用して、一見すると全く違うような分野に挑戦しました。
watashi wa kono gijutsu o ooyoo shite, ikken suruto mattaku chigauyoona bunya ni chosen shimashita.

Applying this technique, I have taken on a challenge in an area that at first glance looks completely different.

Ex. 13-E) 私はこのガンマ線の測定の技術を応用して、一見すると全く違うような分野に挑戦しました。
watashi wa kono ganmasen no sokutei no gijutsu o ooyooshite, ikken suruto mattaku chigauyoona bunya ni choosen shimashita.

Applying this technique for gamma ray measurement, I have taken on a challenge in an area that at first glance looks completely different.

As seen in the examples above, the elaborated versions increased numbers of words, including verbs, because they contained more details while preserving all of the information in the genuine versions (Kim, 2006; O'Donnell, 2009). The fundamental principles of elaboration include adding redundancy and regularity to clarify information. In spoken Japanese, irregular word order also frequently appears. The genuine versions of the listening passages in this study did not include cases of irregular word order, possibly due to the written drafts. When word order was irregular, however, recovering canonical word order is another method that can be used for improving clarity.

4.2.3.3 Modified elaborated version

Modified elaborated input is a new addition to this line of research (Long, 2007, 2015, 2019). These versions were inspired by a line of research that maintains that elaboration involves retention, not removal, of unknown lexical items, in contrast with the simplification of genuine versions. Specifically, the versions were created by preserving almost all the sentences in the elaborated versions, including words and elements of syntactic complexity, but shortening them where necessary. Among the four versions of input employed in this study, the modified elaborated version was hypothesized to be the most effective type of input modification for L2 comprehension and incidental vocabulary learning. The guidelines for creating this version are as follows:

1) The lexical elaboration employed in the elaborated version was used.

2) Long sentences were divided into shorter sentences. When necessary, an inter-sentential connector and/or a pronoun was added at the beginning of a sentence that followed. The examples below contrast with the modified elaborated version and the simplified version of Example 9. The information in the modified elaborated version was exactly the same as in the elaborated version. However, the number of sentences increased as a result of segmenting the elaborated sentences. The number of letters also increased, while the number of S-nodes remained the same in the modified elaborated version (Letters: ME-42 + 34 + 37 + 29 + 39 = 181 S-32 + 30 + 39 = 101; S-nodes: ME-8 S-5). Compared to the number of letters in the genuine version of the example passage,

the modified elaborated version below increased by 131%, and the number of S-nodes doubled.

Ex. 9-ME) 最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告されました。それで、国内での二次感染を防ぐための対策を考える必要がでてきました。わかりやすく言うと、デング熱、ジカ熱などは、熱帯の病気と考えられていました。しかし、最近これらの病気の患者が日本でも報告されています。それで、国内でそのような患者から他の人に移らないようにする必要が出て来ました。

> *saikin, nettai no byooki to kangae rarete ita dengu netsu, jika netsu nado no kanja ga nihon demo hookoku saremashita. sorede, kokunai deno nijikansen o fusegu tameno taisaku o kangaeru hituyoo ga dete kimashita. wakariyasuku iuto, dengu netsu, jika netsu nado wa nettai no byooki to kangaerarete imashita. shikashi, saikin korera no byooki no kanja ga nihon demo hookoku sarete imasu. sorede, kokunai de sono yoona kanja kara hoka no hito ni utsuranai yooni suru hitsuyoo ga dete kimashita.*
>
> Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well. Therefore, strategies that prevent domestic secondary infection are becoming necessary. Simply put, it has been thought that diseases, such as dengue fever and zika fever, are those of tropical areas. However, recently, patients with these diseases have been reported in Japan as well. Therefore, in Japan it is becoming necessary to make sure these diseases do not get passed from these patients to others.

Ex. 9-S) 最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告されました。それで、国内で人から人へ移るのを防ぐための対策を考える必要がでてきました。

> *saikin, nettai no byooki to kangae rarete ita dengu netsu, jika netsu nado no kanja ga nihon demo hookoku saremashita. sorede, kokunai deno hito kara hito e utsuru no wo fusegu tame no taisaku o kangaeru hitsuyoo ga dete kimashita*.
>
> Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well. So, strategies to prevent them getting passed from one person to another domestically are becoming necessary.

4.3 Japanese native speakers' judgments of naturalness of input

To assess the naturalness of the listening materials created for this study, 20 Japanese native speakers were asked to listen to the spoken input and rate it based on their perceptions of how natural it sounded. Each native speaker listened to one version (G, S, E, or ME) of all four talks. Thus, each version was judged by five native speakers. After hearing each talk, they were asked to rate it by selecting one of four choices: unnatural, somewhat unnatural, somewhat natural, natural. As shown in Table 7, among the four versions, the simplified version was judged "unnatural" at the highest rate (10%), while the modified elaborated version was judged "natural" at the highest rate (65%). Overall, however, the created input was judged as natural or somewhat natural by the native speakers, and therefore all input was considered appropriate as materials for the treatment session.

Table 7

Results of Ratings on Naturalness of Input by Japanese Native Speakers

Group *N* = 20	Natural	Somewhat natural	Somewhat unnatural	Unnatural
Genuine	10	9	1	0
n = 5	(50)	(45)	(5)	(0)
Simplified	8	9	1	2
n = 5	(40)	(45)	(5)	(10)
Elaborated	9	8	3	0
n = 5	(45)	(40)	(15)	(0)
Modified Elaborated	13	5	2	0
n = 5	(65)	(25)	(10)	(0)

Note. Percentages appear in parentheses.

4.4 Instrumentation

Comprehension questions on the spoken input and three types of vocabulary tests were administered to the participants. As covariates, one proficiency test and two types of working memory (WM) tests were also administered. Before participants began listening to the spoken input, they were told they would be answering comprehension questions and taking the two WM tests. The vocabulary tests were not announced beforehand; after they completed the comprehension questions, the participants were simply presented with the vocabulary tests and instructed to complete them.[8]

[8] The researcher specifically asked early participants to say nothing about the vocabulary tests to those they introduced for later participation.

4.4.1 Comprehension test

As the participants listened to the four talks, or listening passages, they answered comprehension questions (CQs). The passages were in the auditory mode only, and participants were not allowed to take notes. Due to the length of the passages, each passage was divided into three sections. The length of each section varied from 1 minute 14 seconds to 3 minutes 10 seconds. Immediately after each section, the participant heard two comprehension questions for that section, with four possible choices as answers for each. The comprehension questions also appeared in written form on the screen, so that participants could read the questions. The four-choice answers were not presented on the screen. This procedure was intended to make the participants focus on understanding the content of the listening passage. In addition, after they heard the entire passage, they were asked questions pertaining to the entire talk. The comprehension questions consisted of three types of item: (a) replication items that examined surface comprehension, (b) synthesis items that required the listeners to connect pieces of information in the passage, and (c) inference items that required the listeners to make a deduction about the implications of the passage (Davey, 1988; Yano et al., 1994, p. 205). The replication items were used after the first and second sections. The synthesis items were used after the second and third sections. After the whole passage, the inference items were used. Appendix D shows all CQs used for the four passages.

The comprehension questions had two purposes. First, their use allowed this study to partially replicate Yano et al. (1994), which investigated the relationship between input modification types and comprehensibility, although the current study used aural input instead of reading. Both studies use participants' answers to CQs to examine how modification types facilitate the comprehension of the input. In the current study, the questions were asked in L2 Japanese, to avoid the cost of L1/L2 switching (Meuter & Allport, 1999). The second purpose of the comprehension questions was to determine whether an incidental vocabulary learning (IVL) condition was successfully established. Because participants knew that comprehension questions would be asked after every section, they were expected to focus on the meaning of the passages rather than on novel lexical items. When participants' scores on the comprehension questions were below 50% (18 out of 36 points), the session was discontinued and their scores on the vocabulary tests were excluded from the analyses of IVL. The threshold of half of the maximum score was determined based on the consideration that a score of 50% or higher would demonstrate that (a) the participants' attention was drawn to the meaning of the talks, and (b) their proficiency was sufficiently high to comprehend the talks. Although the participants also took an online proficiency test before the treatment session, it was expected that it would be difficult for the participants to comprehend and to achieve IVL from auditory-only input.[9]

The numbers of comprehension questions and points are provided in Table 8 below. The total possible score for each passage was 9 points, and for all comprehension questions was 36 points.

[9] Sessions were discontinued if the participants were unable to complete all of the procedures within three hours.

Table 8

Total Number of Comprehension Questions and Points in Each Passage (4 versions x 9 points each = 36 points in total; score cut-off: 18 points [50%])

Talk	Section	Replication	Synthesis	Inference	Points
1	1	2			2
	2	1	1		2
	3		2	3	5
		3	3	3	9

4.4.2 Vocabulary learning measures

Based on Kroll and Stewart (1994) and Jiang (2000), the vocabulary learning measures were created to tap into the vocabulary knowledge the participants developed from the limited exposures in the input. Because participants were exposed to the TWs in the L2 alone, their temporary lexical entries for TWs were assumed not to contain clear L1 translations. Rather, the entries were assumed to be supported by the contextual information that the participants had heard along with the novel lexical items. The first test was a form-recognition test, which was intended to test phonological information that the lexical entries were assumed to contain. The second test was a form-meaning recognition test, in which participants heard the TWs in sentences. These sentences were constructed using vocabulary items that were in the listening passages they had heard. Therefore, the sentences were intended to enhance participants' retrieval of the TWs. The third was another form-meaning recognition test, which examined participants' knowledge of the TWs through definitions in Japanese. In previous studies that employed L1 translation in either production or recognition, participants have gained low scores on the translation tests. As Jiang (2000) pointed out, L1 knowledge may be copied in a newly developed lexical entry. However, under incidental learning conditions, it is assumed that the lexical entries are created based on L2 input alone, and that a part of a lemma where L1 semantics would be contained may be missing or contain unclear L2 semantics. The examination of the construct and its validity is beyond the scope of this study. However, the second form-meaning test used the L2 instead of the L1, so that it could examine whether L2 semantics existed in the lexical entries. All of the items were presented through aural input alone. Participants' response times were not recorded.

4.4.2.1 Form-recognition test

The form-recognition test contained 32 aurally presented multiple-choice items. Each item consisted of a single TW along with three fillers, for a total of 128 words (32 TWs and 96 fillers). The fillers were constructed by replacing one or two phonemes in the TW. The replacement was performed at either the first syllable (59.4%), the second syllable (15.6%), or the third syllable or two combined syllables (25.0%). The participants were instructed to select the words they thought they had heard in the listening passages. The entire form-recognition test appears in Appendix E.

4.4.2.2 Form-meaning recognition tests: A sentence test and a definition test

Two types of form-meaning recognition test were administered: one testing for knowledge of the meaning of the TWs in sentences (MST), and the other testing for knowledge of the meaning of the TWs in L2 definitions (MDT). Both were multiple-choice with four options for each item. The tests appear in Appendixes F and G.

The MST consisted of 32 items. In each item, the participant first heard a TW and then listened to four sentences with the TW. Thus, the test contained a total of 128 sentences: 32 targets and 96 fillers. The participants were instructed to select as their answer the sentence in which they thought the TW was used with the same meaning as in one of the listening passages. The test was designed to assess whether participants could recognize the meanings of the TWs when they were presented with comparable contextual information. The sentences were constructed using words from the talks in comparable contexts.

The MDT also consisted of 32 items, one for each TW. In each item, the participant heard one TW and then listened to four definitions in Japanese. They were instructed to select the most appropriate definition based on the listening passages they had heard. The definitions presented in the test were slightly different than the definitions included in the elaborated versions of the talks. The MDT was designed to assess participants' ability to map L2 form to L2 meaning in the absence of context. The test was expected to measure the participants' L2 form-meaning mapping ability rather than their translation ability.

4.4.3 Proficiency measure: Simple Performance-Oriented Test (SPOT)

The proficiency measure test known as the Simple Performance-Oriented Test (SPOT) was developed by Ford-Niwa, Kobayashi, and Yamamoto (1995).[10] It has been widely used in Japanese language placement tests, and has been validated as a concise measure of Japanese proficiency (Hatasa & Tohsaku, 1997; Kobayashi, Sakai, & Ford-

[10] Ford-Niwa et al. (1995) based this test on cloze tests. Its original purpose was to assess Japanese grammatical knowledge.

Niwa, 2007; Suzuki, 2014). Currently available online, it consists of 90 simple sentences (1 point each, total possible score of 90), which are presented both visually (on screen) and auditorily. The visual cue is presented first, to support the participants' parsing of the aural sentence. Each item consists of a sentence with a blank and four choices to fill the blank. The blank was equivalent to one syllable in Japanese, and the four choices showed possible hiragana characters. The missing character is an element of the grammatical structure, such as a case marker. Participants are given four seconds to make each choice, but they can control the pace of the test to some extent because they must use a mouse click to continue onto the next item. The 90 items are divided into three sets. As the test proceeds from one set to another, the difficulty level increases through manipulation of the speed of the aural cue, the salience of the target syllable, aural intelligibility, grammar complexity, and vocabulary level. The test begins with 10 practice items, which participants can work through as many times as they like. The participants can make their answers through either mouse clicks or screen touches, and can view their results at the end of each set. When all 90 items are completed, the results of the entire test appear, and they are downloadable as a pdf file.

Although the pace of the entire test depends on the participant, it took this study's participants approximately 15 minutes. The participants completed the test at home and submitted their results to the researcher. Five participants were unable to take the test due to problems with the website. All other participants were randomly assigned to one of the four treatment conditions, using their proficiency scores to balance the treatment group members' proficiency levels.

4.4.4 WM measures

Malone (2018) suggested that less IVL might occur through aural input alone (i.e., without visual input), due to the greater need for WM resources in aural processing. Findings in the literature, however, are not conclusive regarding aural-only-input in WM resource requirements. Therefore, this study used WM measures to follow up on the suggestion. The literature has suggested that phonological short-term memory (PSTM) measures are relevant for investigating L2 listening comprehension (Baddeley, 2015; Bloomfield et al., 2010; Kormos & Sáfár, 2008; Martin & Ellis, 2012). A task that measures both executive control and PSTM abilities appears to be pertinent. However, a search was unable to find a relevant PSTM measure for Chinese speakers (Z. E. Wen, personal communication, April 9, 2019). Moreover, Linck et al. (2014) reported a significantly larger correlation between L2 proficiency and complex WM span tasks than between L2 proficiency and simple tasks, suggesting a larger role of executive functions than of PSTM in L2 learning. It was also necessary to consider time constraints, as the time needed to listen to the four talks and complete several tests was already considerable. Therefore, this study employed two complex WM measures. Instructions and practice items for both WM tasks were delivered by the researcher on a one-to-one basis in Japanese. The two tasks are described in the following two subsections.

4.4.4.1 Operation Span (OSPAN) task

Complex WM capacities was operationalized using this task. In the OSPAN task, a series of items is presented on a computer, each consisting of an arithmetic equation along with a single alphabet letter (see Appendix H). Participants were instructed to read the equation aloud in their L1 Chinese and determine whether the equation was correct, and to remember the letter. Blocks of from two to five items were randomly presented. At the end of each block of items, participants were asked to recall all the letters they had seen in the block and to write the letters down on an answer sheet. The total number of accurately recalled letters in correct sequence from each block was the participant's test score; the total possible score was 40 points. The participants also received scores on their responses to the veracity of the equations, and this veracity score (also 40 points maximum) was used to decide whether the participant had stayed focused on solving the equations. The cutoff score was 32 points (80%). This test was administered with PychoPy1.85.3.

4.4.4.2 Shapebuilder

Shapebuilder, a visual-spatial task, is another nonverbal complex WM measure (Atkins et al., 2014). Participants track the order and spatial position of a series of colored shapes presented in a 4 x 4 grid on a computer screen (see Appendix I). After seeing such a series, the participants are asked to recall the order, shapes, and colors, and to repeat the exact same series of presentations by dragging the appropriate colored shapes, which are lined up outside of the grid, into the appropriate places in the grid. The length of trials becomes longer, from two to four, as the task progresses. In addition, within each trial, the variety of colors and shapes increases for greater difficulty. Scoring is automatic: (a) 15 points for correct recall of the initial item, (b) 15 points for each additional correct recall of every consecutive item, (c) 5 points for correct recall of color alone with the correct location, and (d) 10 points for correct recall of shape alone with the correct location. There are 25 trials and a total possible score of 3,690 points. Participants were notified of their scores on the computer screen. In this study, the task took the participants approximately 15 minutes to complete. The test was administered with PsychoPy1.85.3.

4.5 Research design and procedure

4.5.1 Design

The study employed a post-test-only, between-subjects control group design. Each participant listened to one type of input modification. The primary independent variable was the type of input modification (four levels: genuine, simplified, elaborated, and modified elaborated). The dependent variables were the scores on the form-recognition vocabulary post-test and the two meaning-recognition vocabulary post-tests (MST and MDT). See Table 9 below for the independent and dependent variables. Comprehension questions were asked for each talk, and the scores were used to examine the relationship between input modification type and comprehensibility. The scores were also used to determine whether participants focused on the content of the talks. The covariates were the scores on the SPOT test of Japanese proficiency and the two WM measures.

4.5.2 Procedure

Data collection was conducted in meeting rooms at participants' universities, a Chinese church, a Japanese church, and a Chinese preparatory school in Tokyo, Japan. The sessions were conducted with one to two participants at a time.

Table 9

List of Variables of the Study

Independent Variable	Dependent Variable 1	Dependent Variable 2	Dependent Variable 3	Dependent Variable 4
Input modification type (G/S/E/ ME)	Scores on comprehension questions (CQ)	Form-recognition vocabulary post-test (FRT)	Form-meaning recognition post-test: sentences (MST)	Form-meaning recognition post-test: L2 definitions (MDT)
	Correct/ Incorrect	Correct/ Incorrect	Correct/ Incorrect	Correct/ Incorrect
Genuine	A-	A	A’	A”
Simplified	B-	B	B’	B”
Elaborated	C-	C	C’	C”
Modified elaborated	D-	D	D’	D”

Covariate: Proficiency, WM (2 tasks)

A treatment session began with the researcher explaining the study and asking the participants to sign a consent form. Each participant then completed a background information sheet. Next, the participants took the online Japanese proficiency test. Based on the results of the proficiency test, the participants were randomly assigned to one of the four treatment groups. In other words, a randomized block design was adopted: Their proficiency was assumed to be a blocking factor due to its relationship with the dependent variables, which were their scores on the comprehension questions and unannounced vocabulary tests (Lomax & Hahs-Vaughn, 2012).

After the researcher provided them with clear instructions for the listening task and showed them the response sheet for comprehension questions, the participants began listening to the listening passages. Each talk was divided into three sections according to the flow of the content. Immediately after each section was heard, participants were led to two comprehension questions (see Appendix J for length of audio files). The questions were presented on the computer screen. Four possible choices for the responses were heard only once. Participants selected the letter that corresponded to their response on their response sheets. At the end of the third section, participants responded to five more comprehension questions, including three inference items. After hearing two listening passages, the participants completed the non-linguistic Shapebuilder task. They then heard the two remaining listening passages.

At that point, the participants took a break while the researcher calculated their scores on the comprehension questions to ascertain their attention to comprehension of the passages. Then, the participants were asked to take the three unannounced vocabulary tests. Each test began with sample test items to ensure the participants understood how to respond before they began the test proper. Following the vocabulary tests, the participants were asked to take the OSPAN test in their L1 Chinese. Finally, the researcher interviewed the participants using a short debriefing questionnaire (Appendix K), which asked about their noticing of the TWs. The purpose of the debriefing session was to gather information that could suggest whether the participants had focused on the content of the talks they had heard, and whether they had noticed the TWs while listening to the talks. The entire session took a minimum of two and a half hours. Table 10 below shows the order and time for the entire treatment session.

Table 10

Timing and Order of Task Administration

Order	Task	Time (in minutes)
1	Background info. sheets & consent form	5
2	SPOT for proficiency	15
3	Talk 1 & CQ	12
4	Talk 2 & CQ	12
5	Shapebuilder	15
6	Talk 3 & CQ	12
7	Talk 4 & CQ	10
	Break	10
8	FRT	10
9	MST	20
10	MDT	15
11	OSPAN	15
12	Debriefing & compensation	15
		166

Chapter 5 Results

5.1 Reliability of measures

Reliability of all measures except the proficiency measure is shown in Table 11 below. Values of Cronbach's alpha were calculated using two packages in R and Excel with split halves reliability also calculated. Both Cronbach's alpha and split halves were examined to compare the values. Overall, they were found to be comparable. The mean reliability coefficients on the three outcome vocabulary tests (FRT, MST, and MDT) were acceptably high (0.80, 0.83, and 0.81 respectively). Coefficients for the WM measures, Shapebuilder (SB) and OSPAN (OS), were 0.76 and 0.77 respectively. These coefficients were comparable to those reported by Malone (2018), which were 0.71 and 0.78. Although the coefficient for comprehension questions (CQ) was lower (0.64), the measure was kept. Long and Ross (1993) stated that a low coefficient on a reading comprehension test (e.g., 0.70) suggests less-homogeneous results for the whole test due to different item types, which require separate underlying comprehension processes. Although the online proficiency test, SPOT, did not provide detailed scores for each item, internal consistency reliability was estimated as 0.70 (Mean: 75.08, *SD*: 6.38, k = 90), using the Kuder-Richardson 21 formula. Thus, the measure was determined to be reasonable.

5.2 Descriptive statistics

Preliminary analyses examining normality assumptions were conducted using IBM SPSS Statistics 25. Table 12 shows the results of descriptive statistics for the comprehension questions and vocabulary post-tests (see also Figures 1–4).

Table 11

List of Measures and Their Reliability Values

	Cronbach's alpha 1 (umx)	Cronbach's alpha 2 (psych)	Cronbach's alpha 3 (Excel)	Split halves (Excel)	Mean
CQ *k* = 36	0.67	0.67	0.66	0.57	0.64
FRT *k* = 32	0.80	0.80	0.81	0.79	0.80
MST *k* = 32	0.84	0.84	0.85	0.80	0.83
MDT *k* = 25	0.82	0.83	0.81	0.81	0.82
SB *k* = 25	0.77	0.77	0.73	0.75	0.76
OS *k* = 40	0.76	0.76	0.78	0.78	0.77
Prof *k* = 90	Due to the use of an online measure, it was impossible to calculate these values.[11]				

All mean scores of the input modification groups were higher than those of the genuine (baseline) group except in one case: the mean score of the simplified group on the FRT. The mean scores of the elaborated group were the highest among the groups across the tests. The study hypothesized that the results of the modified elaborated group would be better than those of both the elaborated and the simplified groups. As shown in the table,

[11] Although the researcher contacted the Center for Distance Learning of Japanese and Japanese Issues, Tsukuba University, which administers the SPOT, the Center provided no information on the reference reliability values of the test.

however, results were mixed; the modified elaborated group has only one mean score, on FRT, higher than the mean score of the simplified group. On the other tests, the simplified group scored higher than the modified elaborated group.

Descriptive statistics for the moderator variables are shown in Table 12 (see also Figures 5–7). On the proficiency test, both elaborated and modified elaborated groups scored higher than the other two groups. On both WM tasks, the elaborated group scored highest of all. According to Scheffe's post-hoc tests, however, statistically significant differences were *not* found between groups in terms of proficiency and WM (Proficiency: $p = 0.861 > 0.05$; WM1: $p = 0.864 > 0.05$; WM2: $p = 0.073 > 0.05$). None of the preliminary data analyses showed statistically significant differences between the four groups.

Table 12

Descriptive Statistics: Scores on CQs and Post-tests

Group [Total Score]	CQ [36]		FRT [32]		MST [32]		MDT [32]	
N = 106	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Genuine	25.37	18–32	20.70	11–29	15.37	7–27	18.00	10–26
n = 27	(4.11)		(4.54)		(4.88)		(4.45)	
Simplified	26.46	19–35	20.62	12–30	16.46	6–30	19.92	6–31
n = 26	(4.87)		(5.08)		(6.50)		(6.45)	
Elaborated	26.80	19–35	23.24	13–31	18.04	6–29	21.64	10–30
n = 25	(4.47)		(4.92)		(6.45)		(5.34)	
Modified Elaborated	26.29	19–32	22.14	14–32	16.14	5–30	19.43	9–30
n = 28	(3.84)		(5.77)		(7.28)		(5.83)	

Note. Standard deviations appear in parentheses.

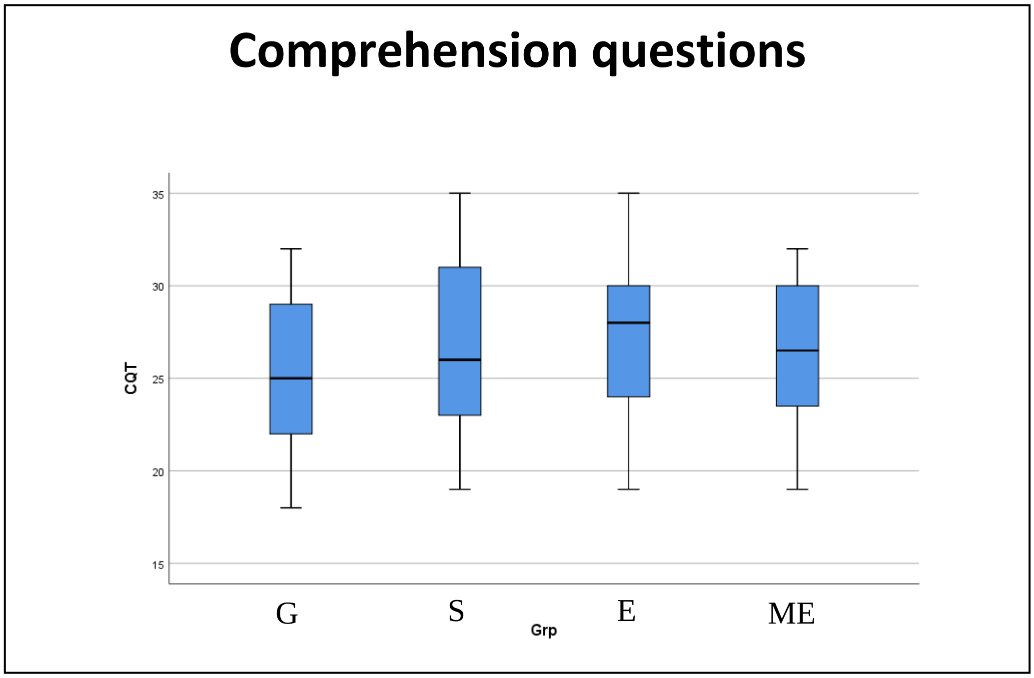


Figure 1. Group mean and score distribution: CQ.

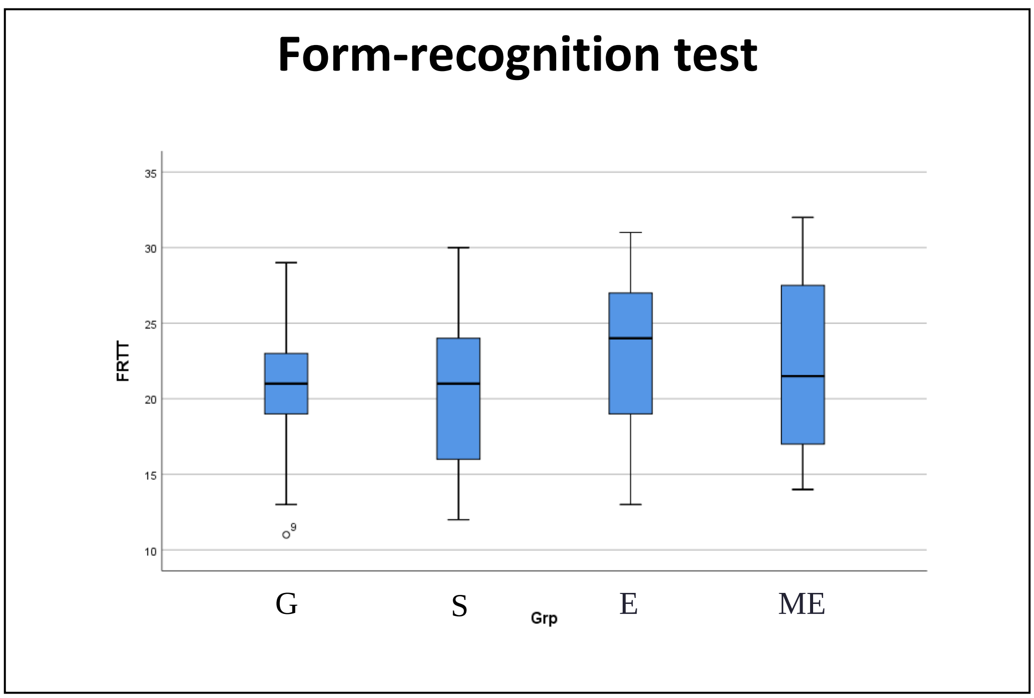


Figure 2. Group mean and score distribution: FRT.

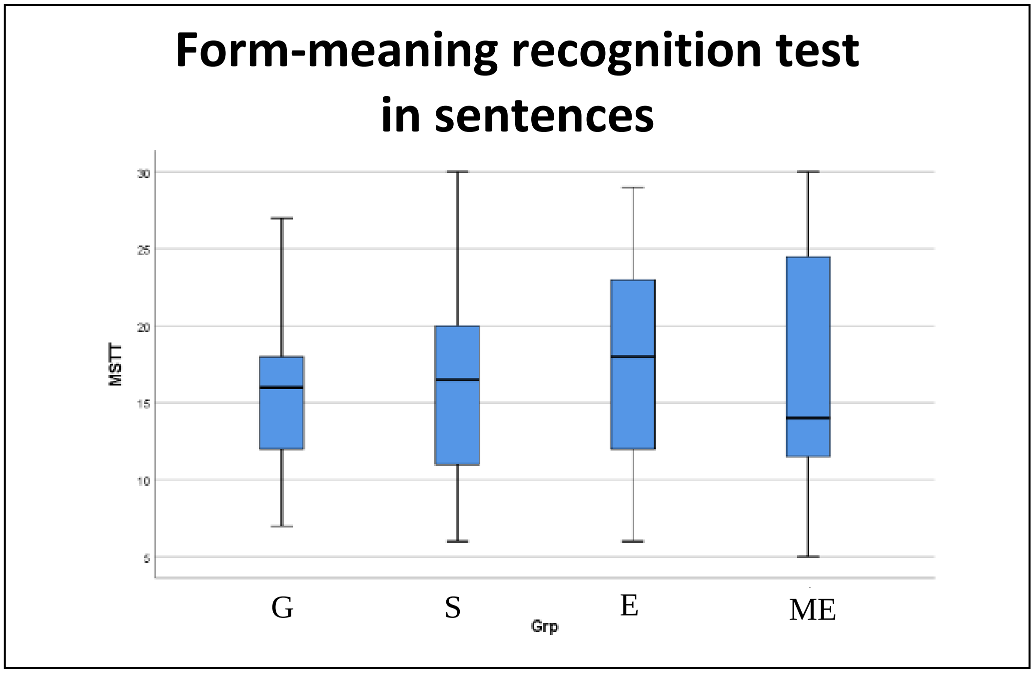


Figure 3. Group mean and score distribution: MST.

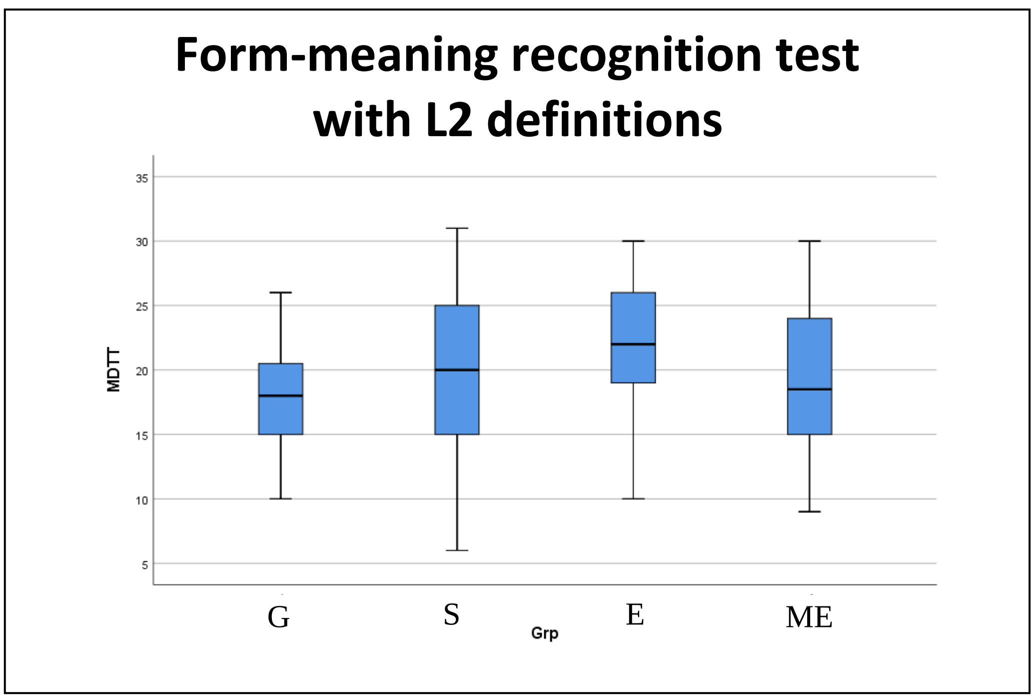


Figure 4. Group mean and score distribution: MDT.

Table 13

Descriptive Statistics: Scores on Proficiency Test and WM Tasks

Group	Prof		WM1		WM2	
[Total score]	[90]		[3690]		[40]	
N = 106	Mean	Range	Mean	Range	Mean	Range
Genuine	74.11	60–84	1619.44	640–2675	27.44	16–37
n = 27	(6.83)		(587.68)		(5.75)	
Simplified	74.96	60–88	1558.65	960–2245	27.88	14–36
n = 26	(7.05)		(422.63)		(5.57)	
Elaborated	75.64	65–88	1666.60	725–2555	31.40	20–39
n = 25	(5.24)		(403.68)		(5.25)	
Modified Elaborated	75.64	61–88	1562.14	985–2570	27.29	17–35
n = 28	(6.43)		(381.72)		(5.74)	

Note. Standard deviations appear between parentheses.

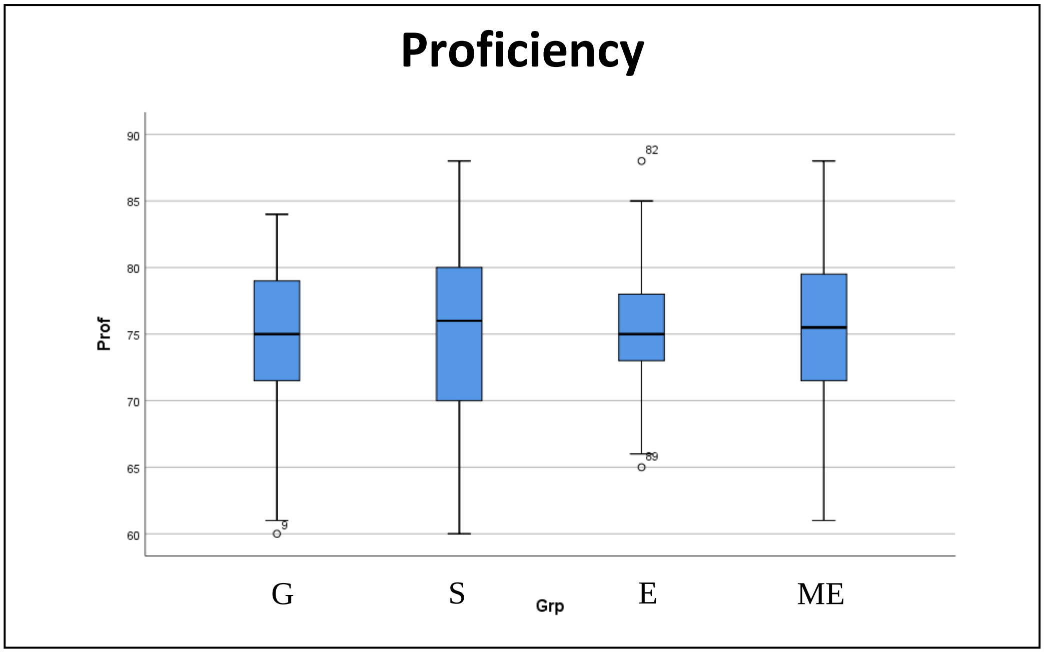


Figure 5. Group mean and score distribution: Proficiency test.

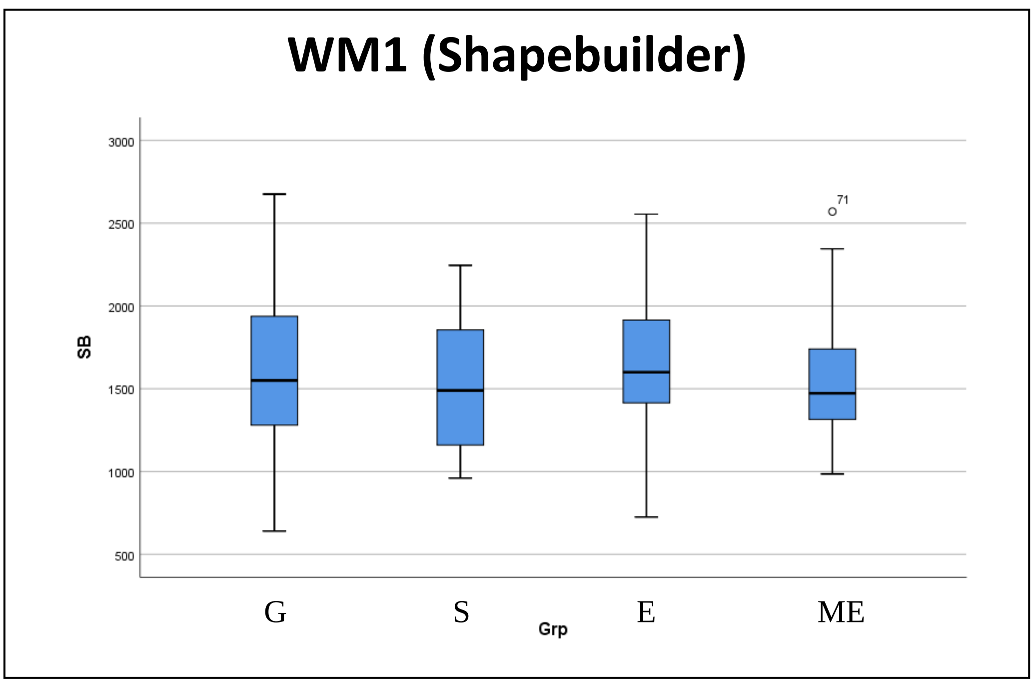


Figure 6. Group mean and score distribution: Shapebuilder (WM1).

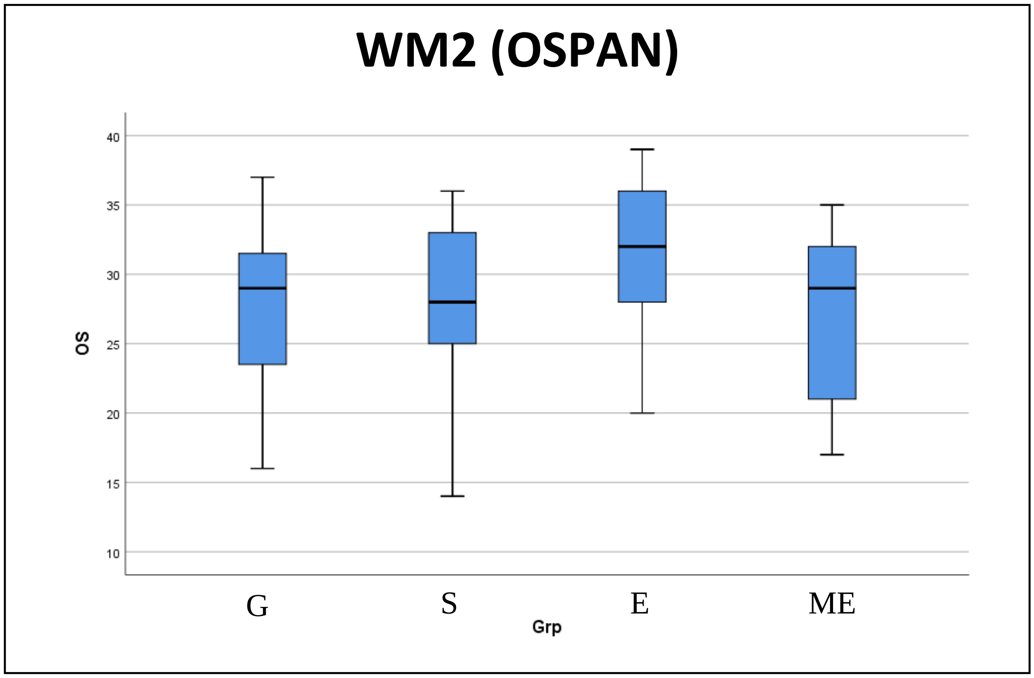


Figure 7. Group mean and score distribution: OSPAN (WM2).

Descriptive statistics of the scores on the three types of comprehension question item by treatment group are shown in Table 14 (see also Figures 8–10). The mean score of the modified elaborated group was highest for replication items, while the mean scores of the elaborated group were highest for synthesis and inference items. The mean scores of the simplified group were the second highest across item types. Results for the modified elaborated group were mixed across item types.

Table 14

Descriptive Statistics of CQ Scores by Item Type

Group [Total score]	Replication [12]		Synthesis [12]		Inference [12]	
N = 106	Mean	Range	Mean	Range	Mean	Range
Genuine	8.19	4–11	8.26	5–11	8.93	5–12
n = 27	(1.84)		(1.75)		(1.84)	
Simplified	8.88	5–12	8.46	5–12	9.12	6–12
n = 26	(1.97)		(1.96)		(1.77)	
Elaborated	8.84	5–12	8.64	5–12	9.32	5–12
n = 25	(1.65)		(1.63)		(2.08)	
Modified Elaborated	9.14	5–12	8.29	4–11	8.86	5–12
n = 28	(1.80)		(1.74)		(1.96)	
All	8.76	4–12	8.41	4–12	9.05	5–12
n = 106	(1.83)		(1.76)		(1.89)	

Note. Standard deviations appear between parentheses.

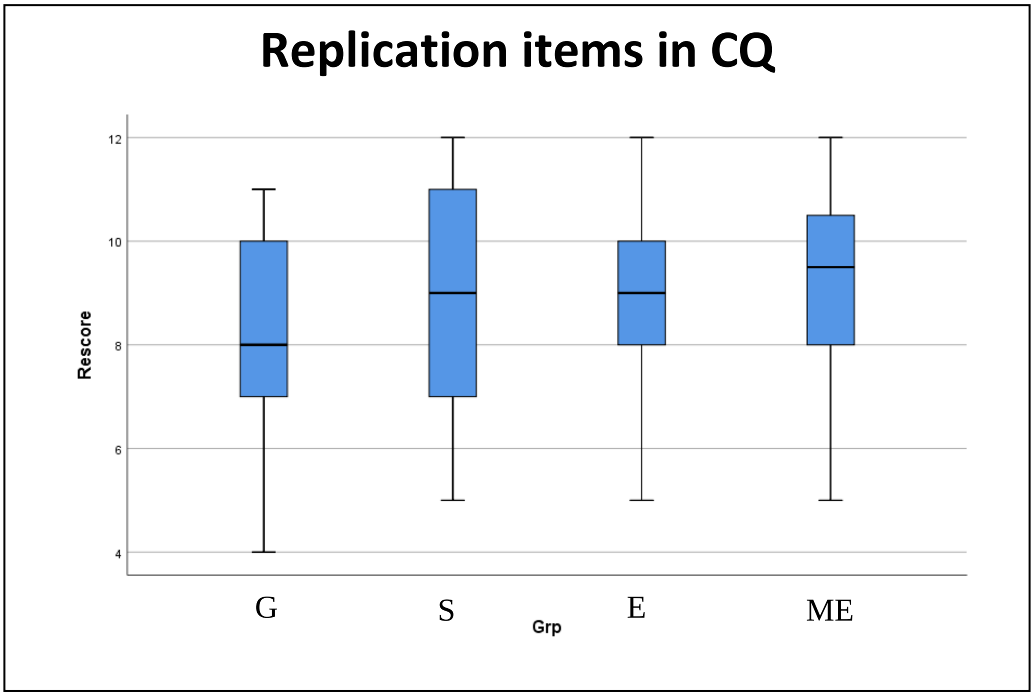


Figure 8. Group mean and score distribution: CQ replication items.

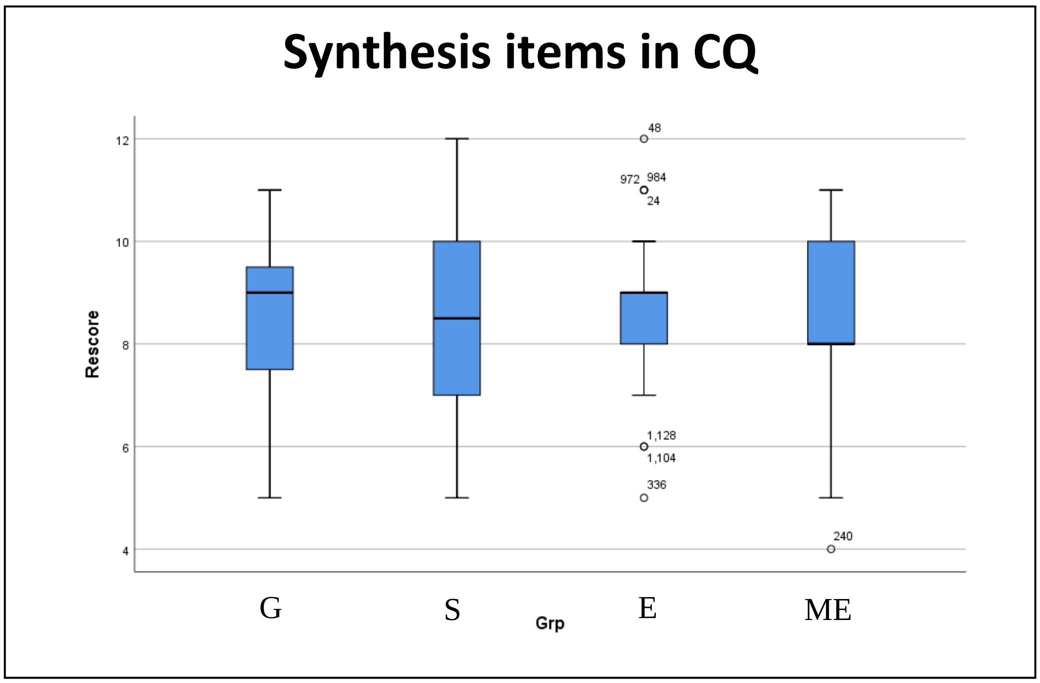


Figure 9. Group mean and score distribution: CQ synthesis items.

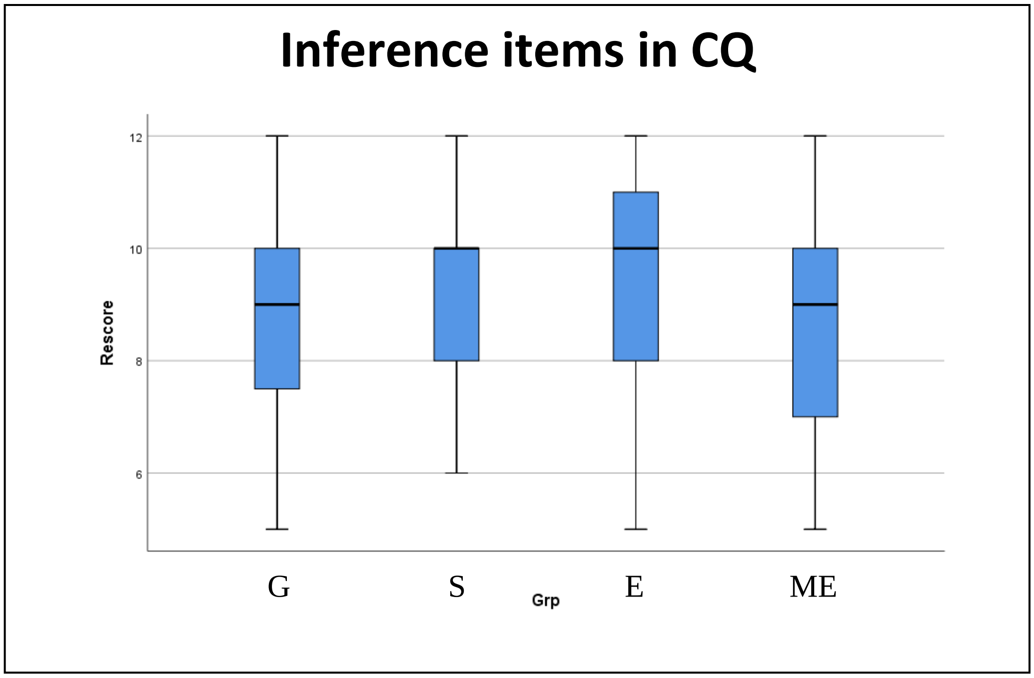


Figure 10. Group mean and score distribution: CQ inference items.

5.3 Normality assumptions

Results of one-sample Kolmogorov-Smirnov (K-S) tests indicated no statistically significant difference from the null hypothesis, which is a normal distribution of data points, on all tests except FRT and MST for the modified elaborated group, and the proficiency test for the elaborated group. Normality test results were below 0.05 (0.034, 0.038, and 0.018 respectively); however, the values of both skewness and kurtosis were within $\pm$1.96 (0.347 and -1.310, 0.418 and -0.917, 0.320 and 0.717 respectively). A visual inspection of frequency histograms, normal Q-Q plots, and box plots showed that the scores of the tests in question were approximately normally distributed. Additionally, all homogeneity of variance values on all test scores revealed no statistically significant difference from the null hypothesis. Therefore, the data were considered to meet the assumptions for further analyses.

Response scores on the three CQ item types were also examined in terms of normality assumptions. The results of one-sample Kolmogorov-Smirnov (K-S) tests showed statistically significant differences for the scores of the modified elaborated group on replication items; the scores of the genuine group, the elaborated group, and the modified elaborated group on synthesis items; and the scores of the genuine and the simplified groups on inference items. However, all values for skewness and kurtosis were within $\pm$1.96 (-0.723 and -0.110; -0.665 and -0.594, -0.115 and -0.416, -0.611 and -0.171; -0.364 and -0.540, -0.468 and -0.907 respectively). All histograms, Q-Q plots, and box plots of all CQ responses by item type were inspected visually. The subsets of the CQ data were judged to meet the assumptions of normality and homogeneity of variance for further analyses.

5.4 Data analysis procedures

Primary analyses of experimental data were conducted using mixed-effects modeling (MEM) for logistic regressions, with binary and continuous outcome variables (Baayen & Milin, 2010; Cunnings, 2012; Hulme et al., 2018; Linck & Cunnings, 2015). The rationale for using mixed-effects models was threefold: (a) they were able to include both categorical and continuous independent and dependent variables; (b) they were able to account for variance of participants and items, as well as their means; and (c) they were robust against violations of sphericity and homoscedasticity (Linck & Cunnings, 2015). In contrast to separate regression models, mixed models are capable of dealing with an entire dataset, including information that may be useful in accounting for focused variable relationships (Winter, 2019).

The lme4 package version 1.1-21 (Bates, Mäachler, Bolker, & Walker, 2016) and R statistical software version 3.5.3 (R Core Team, 2018) were employed for the logistic mixed-effects modeling. The model-building process is shown in Appendix N. With the data sets from both the CQ and vocabulary tests, interaction effects were examined. When they were found, interaction terms were included in the model testing. A forward variable selection method was adopted in which each variable was entered into a model one by one, and the difference made by the variable was examined. This procedure was employed hierarchically. The choice to build the models from simple to complex was made due to the clarity of beginning with the null model. Upon adding a possible explanatory variable, the contribution of the variable in question to the model was shown. The results of the output were interpreted based on the t (z) statistic found in the fixed effects. If the absolute value of the t statistic was greater than or equal to 2.0, the effects were considered significant, and the effects were interpreted as marginally significant if the absolute value was greater than 1.65 (Cook, Pandža, Lancaster, & Gor, 2016; Gelman & Hill, 2007). The models were fit using a maximum likelihood technique. Restricted maximum likelihood (RML) was employed for the estimation method.

Specifically, the process began with spoken input type (group). Within the lme4 package in R, fixed effects were automatically dummy-coded and the genuine group was treated as the baseline. All significant effects in the model were interpreted in regard to the baseline. The dependent variable was accuracy (0, 1), and logistic modeling indicated the probability of response accuracy, given the predictors in the model. Moderator variables were L2 proficiency and WM measures. The purpose of investigating L2 proficiency and WM measures as covariates was to ascertain whether input type made a unique contribution to treatment effects after the individual differences on proficiency and WM measures were accounted for. Proficiency scores and Shapebuilder scores were centered. A composite WM score was created by extracting the means of the Shapebuilder and Operation Span Task scores. The composite score was used for an analysis of CQ scores due to significant correlations of CQ and both WM tasks. The centered Shapebuilder score was used for an analysis of vocabulary test scores due to the lack of a significant correlation between the OSPAN task and the test scores.

Based on the research questions, the analyses consisted of two parts: effects of input type on (a) comprehension and (b) IVL. In the first part, the response variable was comprehension question accuracy, which was binary, with the fixed effect as input modification type. Random effects were participants and question items. All models included random intercepts alone and excluded random slopes due to the single treatment for each participant (Winter, 2019).[12] In conjunction with input type, effects of comprehension question item type were also analyzed. The fixed effect was item type, while the random effects remained the same (participants and items).

Response variables in the second part were the unannounced vocabulary post-test scores, which were also binary, with the fixed effect as input modification type. Random effects were participants and target words (TW).

Additionally, ANCOVAs were performed to compare the models to determine the best fitting models for the data, with the estimation method used as maximum likelihood

[12] It was assumed that within a single treatment, each participant would not vary, and, in addition, neither item nor target word would vary.

(ML). When no increase of the fit to the data was detected by adding random effects to the model, additional effects were removed to maintain parsimony. Because the inclusion of many parameters results in the loss of degrees of freedom, it was determined that the best-fit model should have the minimum possible number of parameters.

Model comparisons were interpreted based on the results of a -2 log likelihood test, shown as Chi-square values, which were ideally smaller for a better fit. These procedures aligned with the evaluations of the logistic regression models: (a) overall model evaluation, (b) statistical tests of individual predictors, (c) goodness-of-fit statistics, and (d) validations of predicted probabilities (Peng, Lee, & Ingersoll, 2002).

Next, using the above scores as continuous outcome variables, one-factor analysis of covariance (ANCOVA) analyses was conducted, and the results were compared with those from the mixed-effects models.

In summary, the overall flow of the analysis procedures was as follows:

(A) Comprehension questions

(1) Logistic regressions for main effects of treatment groups

ANCOVA for main effects of treatment groups

(2) Logistic regressions for main effects of question items

ANCOVA for main effects of question items

(3) Logistic regressions for main effects of question items by group

ANCOVA for main effects of question items by group

(B) Vocabulary post-tests

(1) Logistic regressions for main effects of treatment groups on FRT

ANCOVA for main effects of treatment groups on FRT

(2) Logistic regressions for main effects of treatment groups on MST

ANCOVA for main effects of treatment groups on MST

(3) Logistic regressions for main effects of treatment groups on MDT

ANCOVA for main effects of treatment groups on MDT

What follows is a report of these analyses' results, which include the best-fitting models. Because the overall results of the ANCOVAs were found to be aligned with those of the logistic mixed-effects modelings, only the MEM results are reported in this chapter. Results from the ANCOVAs can be found in Appendix L.

5.5 Group comparisons: Comprehension question scores

5.5.1 Logistic regressions for main effects of input type

Before performing logistic analyses, interaction effects between variables were examined, including random intercepts for both participants and question items. No interaction effects were found.

Following the analyses of interaction effects, the logistic MEM model-fitting procedures were performed. The explanatory variables were input type, proficiency, and composite WM. The response variable was CQ responses. As a result, the maximum model, including both proficiency and composite WM, was determined to be the best fitting model (AIC = 4020.2, BIC = 4070.2, -2 log likelihood = -2002.1, deviance = 4004.2, Chi-square = 41.57). Results showed no significant effect of input type on responses to comprehension questions (all *ps* > 0.10; see Table 15). Conversely, the covariates in the model (proficiency and composite WM) showed a statistically significant relationship with accuracy within the baseline (genuine) condition (z = 5.66, p

< 0.001; $z = 2.67$, $p < 0.01$). This differs from earlier findings where simplified and elaborated versions resulted in significantly better outcomes than the genuine version (e.g., Yano et al., 1994). Descriptive statistics showed higher mean scores for all modified input versions over the genuine version. Relationships between total scores and moderator variables are shown in Figures 11 and 12. When controlling for WM, scores on the proficiency test in the four input groups showed positive linear relationships with scores on comprehension questions (see Figure 11). Although no interactions were found, the relationships between groups and WM scores suggest that both simplified and modified elaborated groups required higher WM capacities to gain better comprehension outcomes when controlling for proficiency. The relationships between WM and the outcomes of the two groups whose input had shorter sentences (simplified and modified elaborated) were found to be comparable, as shown in the profile plot of Figure 12.

Table 15

Results of Logistic MEM for Input Type on Comprehension Question Scores

Best Fitting Model [CQ binary response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item)]

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	1.14	0.19	5.98	0.000***
Simplified	0.16	0.16	0.99	0.32
Elaborated	0.10	0.17	0.60	0.55
Modified Elaborated	0.08	0.16	0.53	0.60
Proficiency	0.34	0.06	5.66	0.000***
Composite WM	0.16	0.06	2.67	0.008**
Random effects	**Variance**	**SD**		
Intercept |Participant	0.18	0.42		
Intercept |Item	0.84	0.92		

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

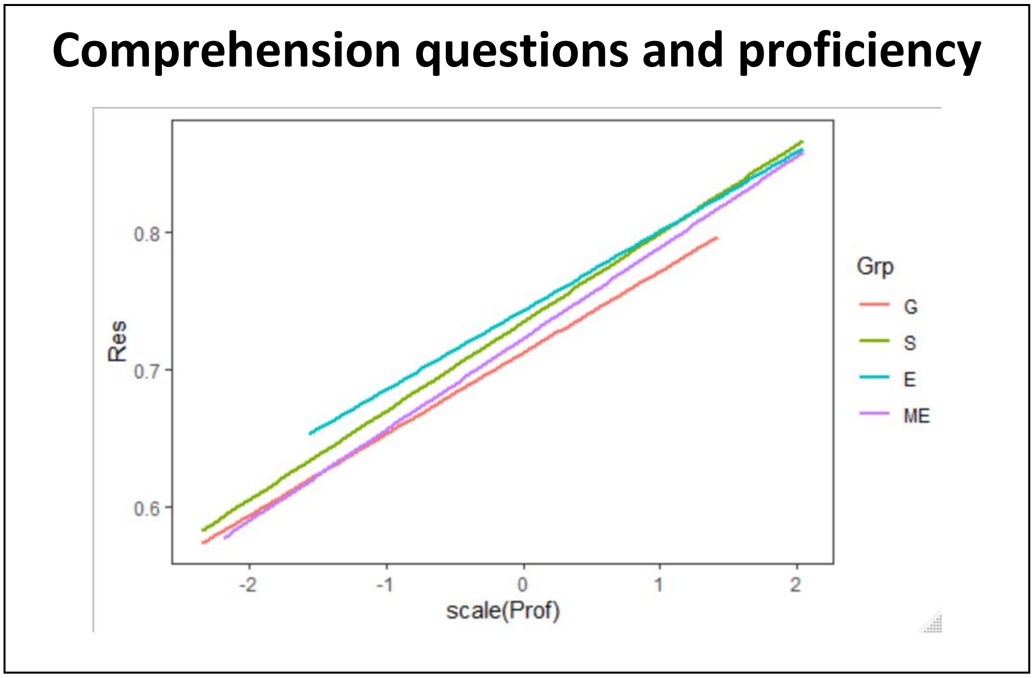


Figure 11. Profile plot of relationships between scores on CQ and proficiency by group.

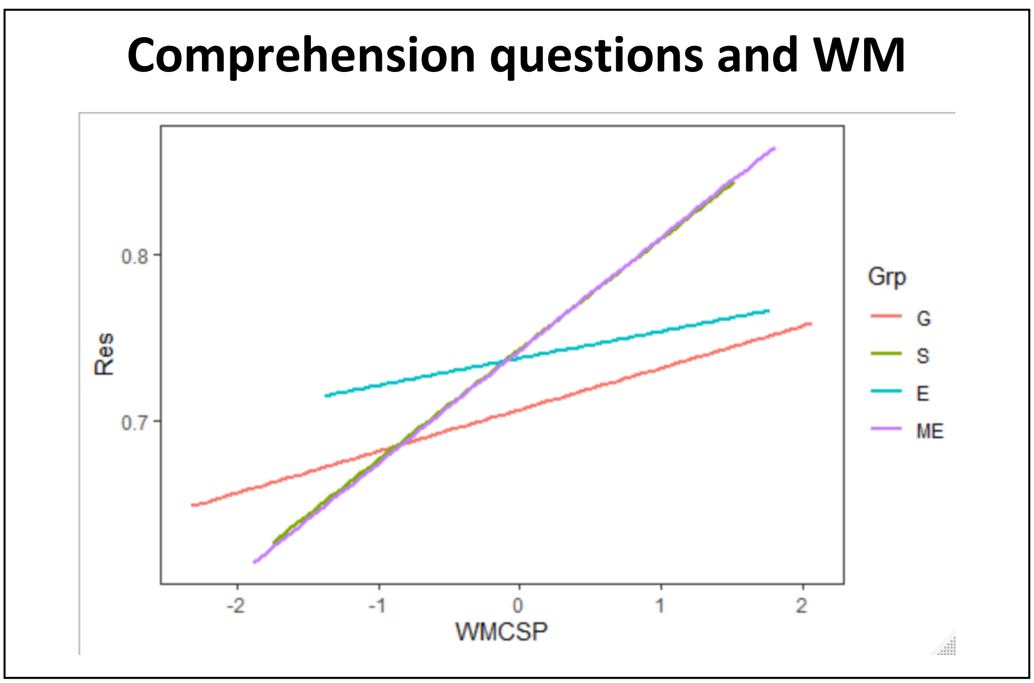


Figure 12. Profile plot of relationships between scores on CQ and composite WM by group.

5.5.2 Logistic regressions for main effects of CQ item types

A logistic MEM was performed with regard to CQ item types. The explanatory variables were CQ item type, proficiency, and composite WM. The response variable was CQ responses. The results showed no statistically significant difference between question items in the best fitting model (AIC = 4018.8, BIC = 4062.5, -2 log likelihood = -2002.4, deviance = 4004.8, Chi-square = 42.36; see Table 16). Fixed effects of both synthesis and inference items were not statistically significantly different from those of replication items. Conversely, both the covariates in the model (proficiency and composite WM) showed a statistically significant relationship with accuracy within the baseline (genuine) condition ($z = 5.72$, $p < 0.001$; $z = 2.77$, $p < 0.01$). As shown in Figures 13–14, the relationships between question item type and the covariates were comparable, with higher scores on inference items than synthesis items across proficiency and WM capacities. Scores on replication items landed between those on inference and synthesis items.

Table 16

Results of Logistic MEM for Item Type on CQs

Best Fitting Model [CQ binary response ~ Item Type + centered Prof + composite WM + (1|Prsn) + (1|Item)]

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Replication)	1.28	0.28	4.62	0.000***
Synthesis	-0.19	0.39	-0.49	0.62
Inference	0.03	0.39	0.07	0.94
Proficiency	0.34	0.06	5.72	0.000***
Composite WM	0.16	0.06	2.77	0.006**

Random effects	Variance	SD
Intercept |Participant	0.18	0.43
Intercept |Item	0.83	0.91

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

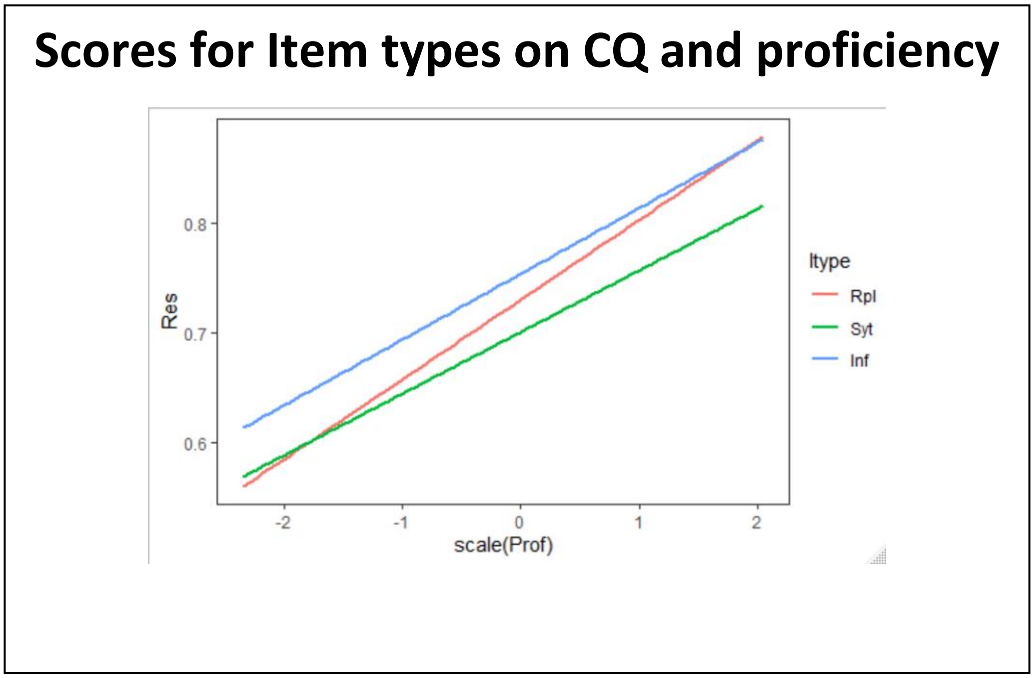


Figure 13. Profile plot of relationships between scores on three CQ item types and proficiency.

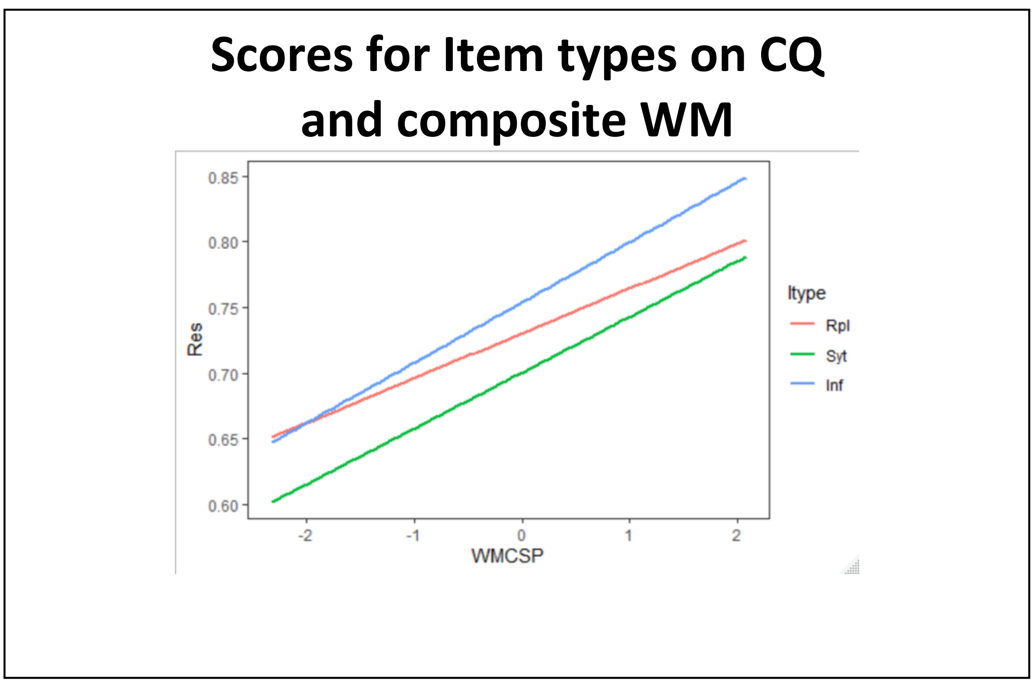


Figure 14. Profile plot of relationships between scores on three CQ item types and composite WM.

To further investigate relationships between item type and input type, a logistic MEM was performed for item type by input group.

5.5.3 Logistic regressions for main effects of CQ item type by group

5.5.3.1 Replication items

The relationships between input type and replication items were the next to be investigated. The response variable was responses to CQ replication items, which were binary (0, 1). The explanatory variables were input type, proficiency, and composite WM. The logistic MEM results showed the maximum model to be the best fitting model (AIC = 1308.3, BIC = 1349.5, -2 log likelihood = -646.16, deviance = 1292.3, Chi-square = 29.38). As Table 17 shows, the modified elaborated group showed a weak effect (b = 0.41, SE = 0.22, p = 0.06). As to covariates, only proficiency showed a significant effect (b = 0.41, SE = 0.08, p = 0.00). To ascertain the results from MEM, the results of an ANCOVA for replication items were also examined. They aligned with the MEM results, showing a trend toward significance with the modified elaborated group (b = 0.79, SE = 0.44, p = 0.07) and a statistically significant difference by proficiency (b = 0.79, SE = 0.16, p = 0.00). These findings are contrary to those of Yano et al. (1994), in which the simplified group performed best with replication items.

Table 17

Results of Logistic MEM for Input Type on CQ Replication Items

Best Fitting Model [CQ binary Replication response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item)]

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	1.07	0.34	3.15	0.001**
Simplified	0.31	0.22	1.41	0.16
Elaborated	0.19	0.22	0.84	0.40
Modified Elaborated	0.41	0.22	1.88	0.06†
Proficiency	0.41	0.08	4.93	0.000***
Composite WM	0.13	0.08	1.50	0.134
Random effects	**Variance**	**SD**		
Intercept |Participant	0.14	0.37		
Intercept |Item	1.08	1.04		

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

5.5.3.2 Synthesis items

The synthesis items in the CQs were also investigated in relation to input type. The response variable was responses to CQ synthesis items, which were binary (0, 1). The explanatory variables were input type, proficiency, and composite WM. By examining the interaction effects, composite WM showed interaction effects with the simplified group ($b = 0.59$, $SE = 0.20$, $z = 2.90$, $p = 0.004$) and the modified elaborated group (*b* = 0.44, *SE* = 0.20, *z* = 2.18, *p* = 0.029). Thus, the interaction term was included in the model testing.

A logistic MEM was then employed for the CQ synthesis items, and the maximum model with the interaction term, Input type x Composite WM, was found to be the best fitting model (AIC =1405.0, BIC = 1461.6, -2 log likelihood = -691.48, deviance = 1383.0, Chi-square = 30.69). There were no fixed effects for group. The interaction effect between simplified group and WM showed a significant effect on responses to synthesis items (*b* = 0.52, *SE* = 0.20, *p* = 0.008). There was also a weak interaction effect between modified elaborated group and WM (*b* = 0.37, *SE* = 0.20, *p* = 0.06). Sentence lengths were shorter in these two input types than in the other two types, and these findings suggest that complex WM could play a role in integrating information available from shorter utterances. Additionally, proficiency showed a significant relationship with accuracy on synthesis items (*b* = 0.24, *SE* = 0.07, *p* = 0.00). However, composite WM showed no effects as a moderator variable, as shown in Table 18.

Table 18

Results of Logistic MEM for Input Type on CQ Synthesis Items

*Best Fitting Model [CQ binary Synthesis response ~ Input Type + centered Prof + composite WM + Input type*composite WM + (1|Prsn) + (1|Item)]*

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	1.02	0.32	3.16	0.001**
Simplified	0.17	0.20	0.85	0.39
Elaborated	0.12	0.21	0.60	0.55
Modified Elaborated	0.04	0.20	0.20	0.84
Proficiency	0.24	0.07	3.31	0.000***
Composite WM	-0.02	0.11	-0.20	0.840
Simplified x Composite WM	0.52	0.20	2.65	0.008**
Elaborated x Composite WM	0.06	0.20	0.31	0.756
Modified x Composite WM Elaborated	0.37	0.20	1.88	0.060†

Random effects	Variance	SD
Intercept |Participant	0.03	0.18
Intercept |Item	1.01	1.00

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

5.5.3.3 Inference items

Results from the inference items in the CQs were also investigated in relation to input type. The response variable was binary responses to CQ inference items. The explanatory variables were input type, proficiency, and composite WM.

A logistic MEM was then conducted for inference items, and the maximum model with the interaction term, Input type x Proficiency x Composite WM, was found to be the best-fit model (AIC =1362.0, BIC = 1454.7, -2 log likelihood = -663.01, deviance = 1326.0, Chi-square = 29.23). Although there were no fixed effects for group, there was a significant effect of proficiency ($b = 0.47$, $SE = 0.15$, $p = 0.003$). A three-way interaction showed a weak trend with the modified elaborated group alone ($b = -0.37$, $SE = 0.22$, $p = 0.095$). These results, shown in Table 19, do not align with the finding of Yano et al. (1994) that the elaborated group scored significantly higher than the genuine and simplified groups on inference items.

Figures 15 and 16 show the response results for each of the three CQ item types in relation to proficiency by input group. When WM is controlled, the elaborated group's scores on the inference items were higher than on the other items. The modified elaborated group scored higher on replication items when proficiency was higher. When the analysis controls for proficiency, higher WM scores are correlated with higher scores on all CQ item types by the participants in the modified elaborated group. In contrast, the WM capacities of participants in the elaborated input group was not related to their scores on either replication or synthesis item types, but was related to their scores on inference items. It is noteworthy that the scores on the synthesis items in the genuine input group appear to be less affected by both moderator variables. This might suggest that the participants were accustomed to synthesizing information to obtain the gist of input in their daily lives.

Figures 17 and 18 are the profile plots of CQ responses and moderator variables by item type. When the analyses control for WM, the more proficient participants in the modified elaborated and simplified groups scored higher on replication items. With synthesis items, participants with lower proficiency in the simplified and the modified elaborated groups scored lower than those in the genuine group. In contrast, participants with higher proficiency in the same groups scored higher than those in the genuine group. With inference items, genuine input is most affected by proficiency, which differs from the results of the other item types. When proficiency is controlled, the plot of the inference items differs from those of the replication and the synthesis items. With synthesis items, both the elaborated and genuine input groups did not require WM resources. In contrast, the modified elaborated and simplified input groups did require WM resources. The statistically marginally significant and strongly significant interaction effects found in the MEM align with the plots. With inference items, however, no such difference was found.

Table 19

Results of Logistic MEM for Input Type on CQ Inference Items

*Best Fitting Model [CQ binary Inference response ~ Input Type + centered Prof + composite WM + Input Type*centered Prof*composite WM + (1|Prsn) + (1|Item)]*

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	1.32	0.25	5.32	0.000***
Simplified	-0.05	0.24	-0.21	0.834
Elaborated	-0.003	0.25	-0.01	0.991
Modified Elaborated	0.008	0.25	0.03	0.972
Proficiency	0.47	0.15	3.02	0.003**
Composite WM	0.20	0.15	1.34	0.179
Simplified x Proficiency	-0.12	0.23	-0.51	0.613
Elaborated x Proficiency	-0.06	0.27	-0.22	0.826
Modified x Proficiency Elaborated	-0.34	0.23	-1.49	0.137
Simplified x Composite WM	0.01	0.24	0.05	0.962
Elaborated x Composite WM	0.05	0.25	0.22	0.827
Modified x Composite WM Elaborated	0.10	0.25	0.39	0.695
Proficiency x Composite WM	0.06	0.12	0.55	0.582
Simp x Prof x Composite WM	0.21	0.22	0.95	0.345
Elab x Prof x Composite WM	-0.17	0.32	-0.52	0.603
Modified x Prof x Composite WM Elaborated	-0.37	0.22	-1.67	0.095†

Random effects	Variance	SD
Intercept |Participant	0.18	0.43
Intercept |Item	0.39	0.63

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

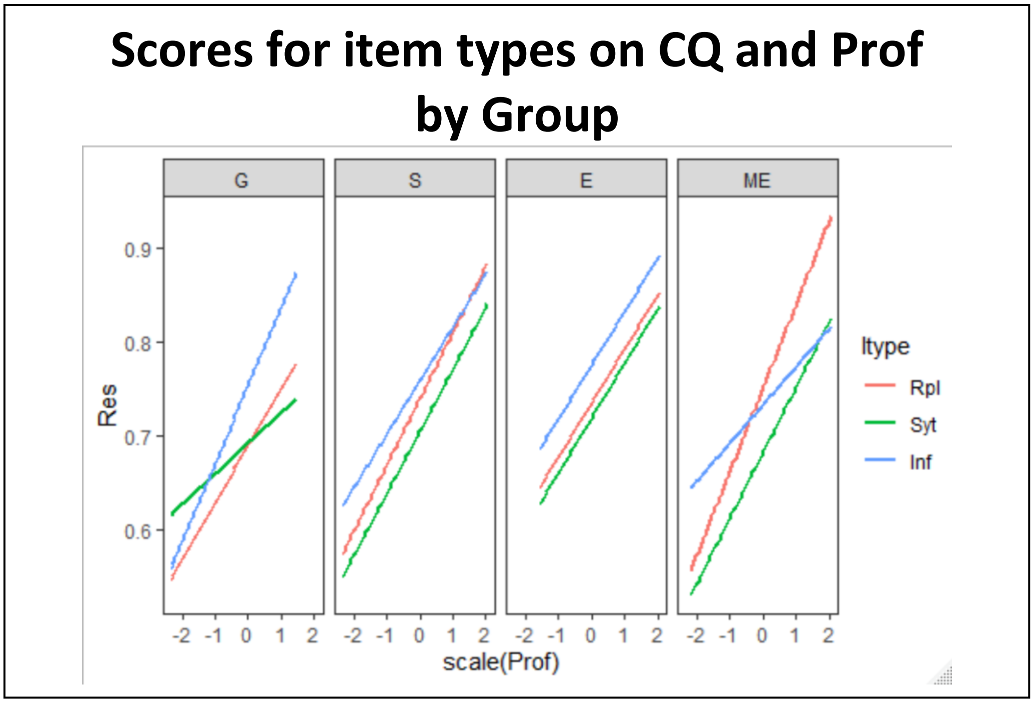


Figure 15. Profile plots of relationships between scores on three CQ item types and proficiency by group.

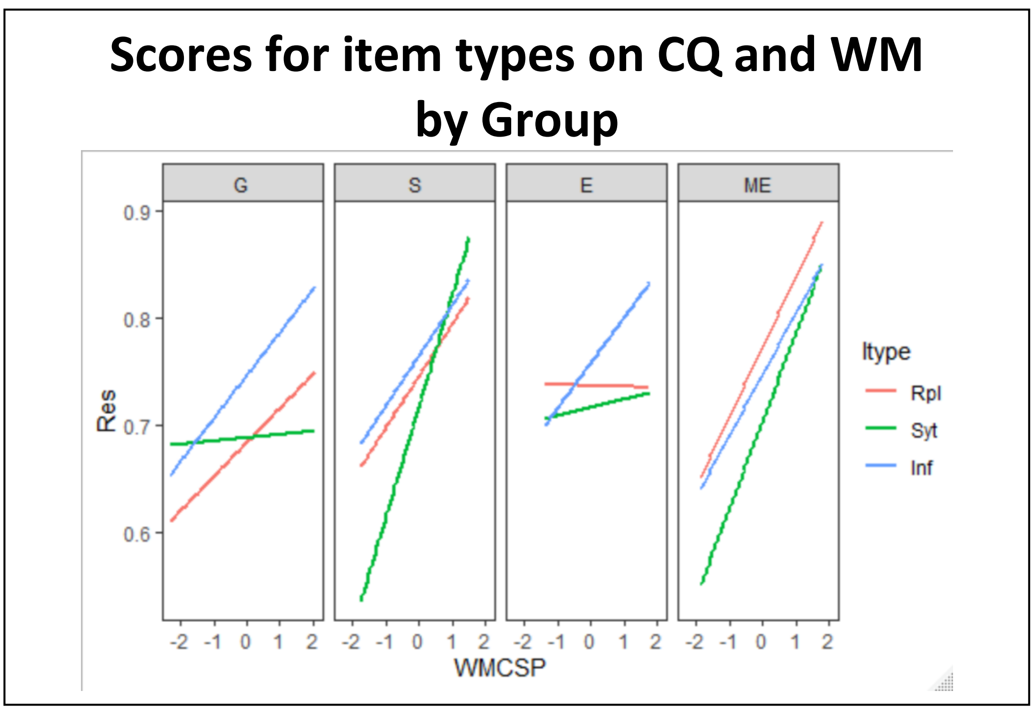


Figure 16. Profile plots of relationships between scores on three CQ item types and composite WM by group.

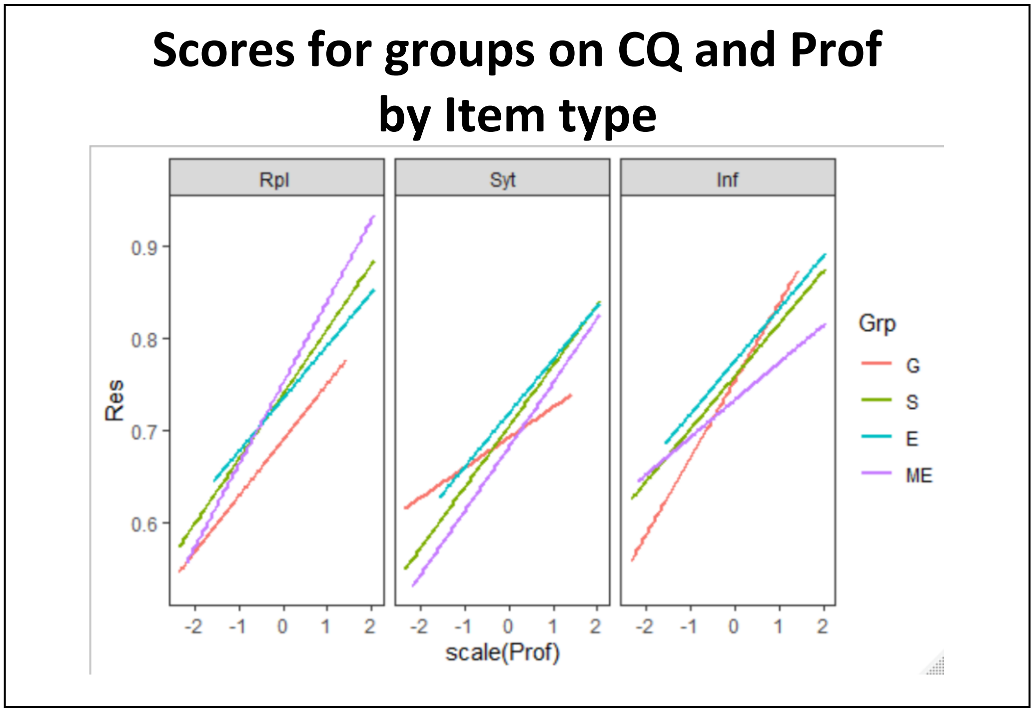


Figure 17. Profile plots of relationships between group scores on CQ and proficiency by item type.

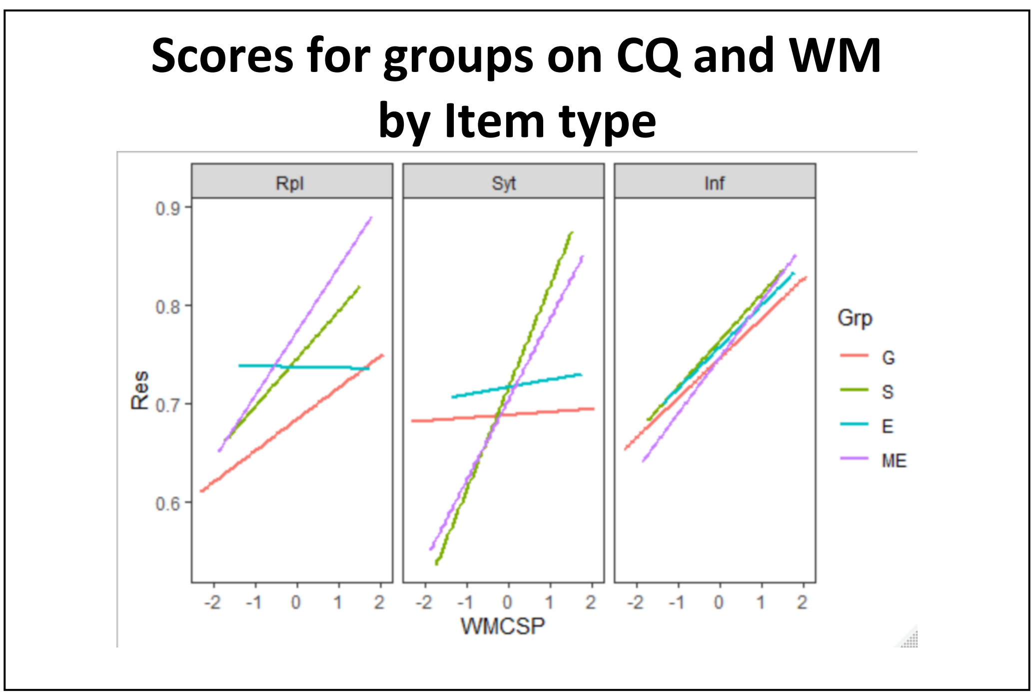


Figure 18. Profile plots of relationships between group scores on CQ and composite WM by item type.

5.6 Group comparisons: IVL

5.6.1 Logistic regressions for main effects of input type on FRT

As with the previous analyses for comprehension question responses, the lme4 package in R was used for MEM model-fitting procedures. The fixed effect of input type was automatically dummy-coded for each group. Statistical significance was interpreted in relation to the baseline provided by the scores for the genuine input group. In the logistic MEM, accuracy (0, 1) on vocabulary tests was the outcome variable, and a model was selected as the best fit based on the probability of response accuracy, given the predictors in the model.

For predicting FRT scores as the response variable, the maximum model with both covariates (proficiency and WM) and the interaction term, Input type x WM, showed the best fit, indicating the smallest -2 log likelihood (AIC = 3524.2, BIC = 3591.6, -2 log likelihood = -1751.1, deviance = 3502.2, Chi-square = 53.72). As a fixed effect, elaborated group had a marginally statistically significant effect on FRT scores (b = 0.41, SE = 0.22, p = 0.063). For the moderator variables, only proficiency showed a significant effect on outcome scores (b = 0.53, SE = 0.08, p = 0.000). A statistically significant interaction effect was found between WM and modified elaborated group, which suggests that the participants whose WM capacities were larger in the modified elaborated group earned higher scores on the FRT (b = 0.55, SE = 0.23, p = 0.018). The results are shown in Table 20.

Table 20

Results of Logistic MEM for Input Type on FRT

*Best Fitting Model [FRT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB) + (1|Prsn) + (1|TW)]*

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	0.94	0.26	3.57	0.000***
Simplified	-0.05	0.22	-0.23	0.820
Elaborated	0.41	0.22	1.86	0.063†
Modified Elaborated	0.29	0.22	1.34	0.180
Proficiency	0.53	0.08	6.52	0.000***
WM	0.04	0.12	0.34	0.733
Simplified x WM	0.12	0.20	0.58	0.562
Elaborated x WM	0.13	0.22	0.59	0.554
Modified x WM Elaborated	0.55	0.23	2.37	0.018*

Random effects	Variance	SD
Intercept |Participant	0.42	0.65
Intercept |Item	1.47	1.21

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

The profile plots in Figures 19 and 20 show that the elaborated group performed better than the genuine group on FRT, showing a weak trend of an elaboration effect. Among the three input modification groups, the simplified group scored lower overall than the other groups on the FRT. This finding cannot be directly compared to results from other studies due to differences in conditions (e.g., proficiency level) and outcome variables (e.g., reading comprehension scores vs. unannounced vocabulary post-test scores). However, it does indicate that both types of elaborated input facilitated more learning of TW forms from listening than did the simplified input, when controlling for WM. These results also make it clear that proficiency played a strong role in the achievement of higher scores.

With regard to the relationships between input type and WM, the plot revealed unexpected results. On the FRT, the modified elaborated input was more effective than the treatments of the other groups (S, E, G) when WM capacities were larger. The finding that the outcomes for the elaborated input group were less affected by WM capacities than the outcomes for the modified elaborated input group was unexpected. Additionally, the plot revealed that simplified input called for more WM capacities than genuine input. The finding of a significant interaction effect between modified elaborated group and WM on FRT outcomes aligns with the finding of a marginally significant interaction between modified elaborated group and synthesis comprehension question scores, as reported above. It is noteworthy that the interaction was found even in TW form recognition, which suggests that WM could be centrally involved in integrating word forms heard in shorter sentences.

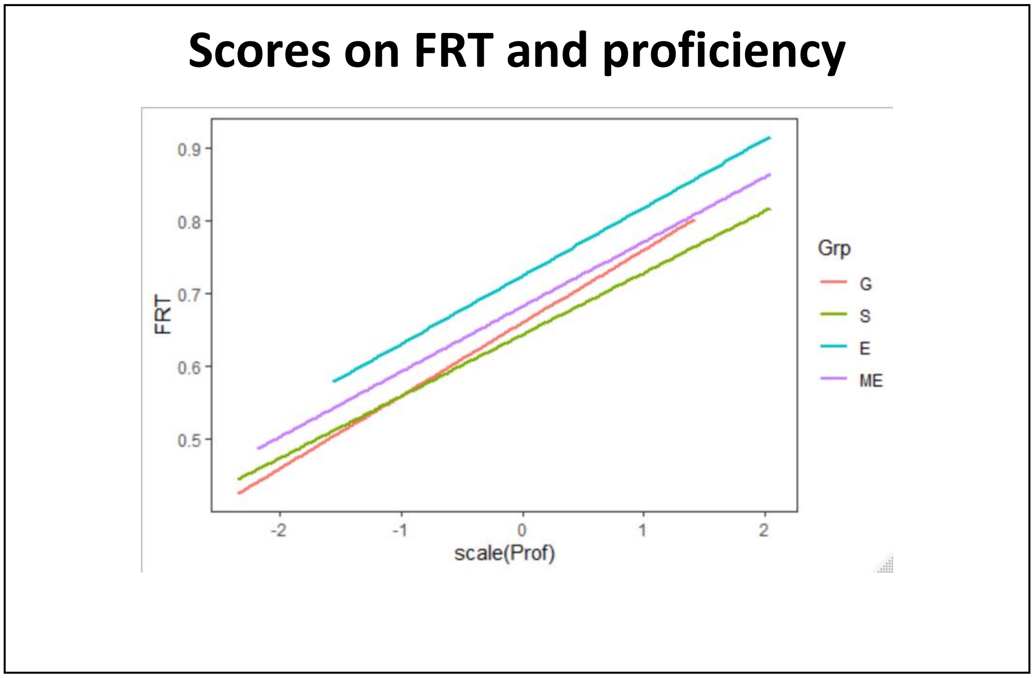


Figure 19. Profile plot of relationships between group scores on FRT and proficiency.

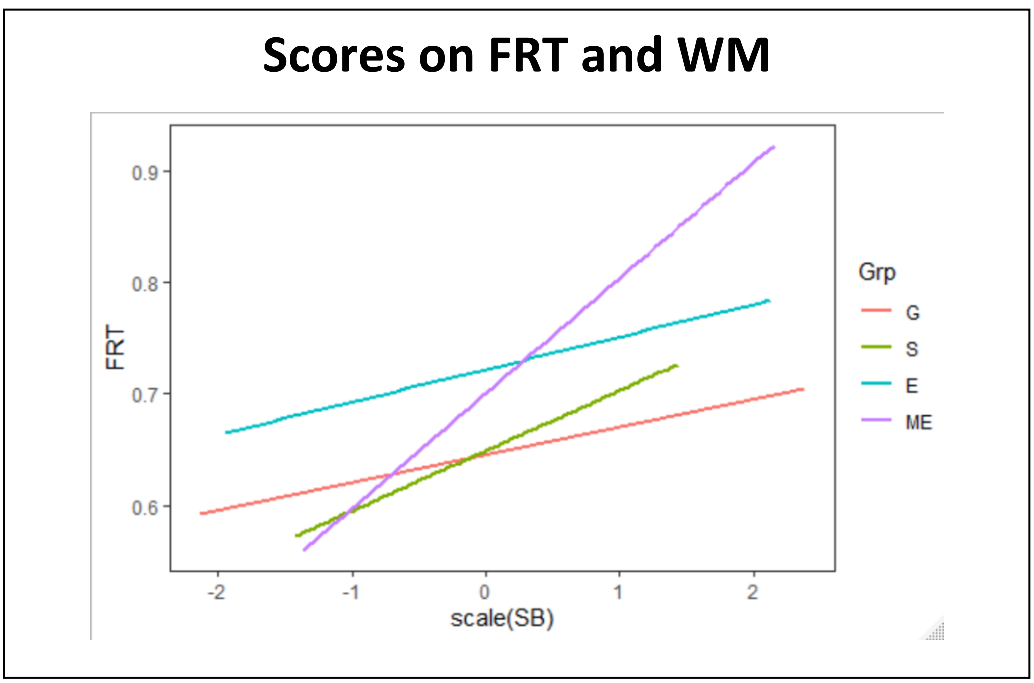


Figure 20. Profile plot of relationships between group scores on FRT and WM.

5.6.2 Logistic regressions for main effects of input type on MST

An examination of interaction effects between input type, MST scores, and proficiency revealed a three-way interaction effect. Therefore, in the subsequent logistic MEM, a three-way interaction term was added in the maximum model. The explanatory variables were input type, proficiency, and WM. The response variable was MST binary scores. As a result, the maximum model was the best-fit (AIC = 3955.9, BIC = 4066.2, -2 log likelihood = -1959.9, deviance = 3919.9, Chi-square = 68.36), as shown in Table 21. The fixed effect for elaborated group showed a weak significant effect (b = 0.41, SE = 0.22, p = 0.055). The interaction effect of modified elaborated group with WM was found to be marginally statistically significant (b = 0.50, SE = 0.27, p = 0.069). The three-way covariate interactions with simplified group suggested a trend for weak effects (b = 0.35, SE = 0.19, p = 0.069). Proficiency showed a statistically significant relationship with accuracy within the baseline (genuine) condition on MST scores, whereas WM did not.

In the profile plots (Figure 21), the strong significant effect of proficiency on MST scores is shown when controlling for WM. Among input modification groups, the elaborated group consistently scored higher than the other groups. The modified elaborated group scored lower than the other input modification groups on MST. Scores of both modified elaborated and simplified groups were lower than those of the genuine group when participants' proficiency test scores were low. However, when controlling for proficiency, the analysis showed that participants with larger WM capacities in the modified elaborated group *outperformed* the other groups, as shown in Figure 22. In contrast, the elaborated group showed almost no effects of WM. At face value, this seems counterintuitive, given the length and the complexity of sentences in the elaborated versions. One possibility is that input elaboration, such as redundancy and appositive cues for TWs, is beneficial for cognitive processing, as it requires fewer WM resources. The role of WM in storing information could be less important when listening to talks in which information is repeated. On the other hand, the modified elaborated version clearly created a larger cognitive load due to the listeners' need to integrate information from segmented sentences.

Table 21

Results of Logistic MEM for Input Type on MST

*Best Fitting Model [MST binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB)*centered Prof + (1|Prsn) + (1|TW)]*

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	-0.09	0.23	-0.38	0.701
Simplified	0.09	0.22	0.44	0.663
Elaborated	0.41	0.22	1.92	0.055†
Modified Elaborated	0.14	0.22	0.61	0.542
Proficiency	0.40	0.15	2.75	0.006**
WM	0.13	0.12	1.07	0.286
Simplified x WM	0.08	0.21	0.39	0.697
Elaborated x WM	-0.10	0.21	-0.45	0.651
Modified x WM Elaborated	0.50	0.27	1.82	0.069†
Simplified x Proficiency	0.33	0.21	1.58	0.114
Elaborated x Proficiency	0.31	0.24	1.28	0.202
Modified x Proficiency Elaborated	0.12	0.22	0.53	0.593
Proficiency x WM	0.09	0.10	0.88	0.379
Simplified x Proficiency x WM	0.35	0.19	1.82	0.069†
Elaborated x Proficiency x WM	-0.23	0.36	-0.64	0.520
Modified x Proficiency x WM Elaborated	-0.11	0.21	-0.51	0.614

Random effects	Variance	SD
Intercept |Participant	0.41	0.64
Intercept |TW	0.90	0.95

Note. Significant at $p < 0.001$ ‘***’, $p < 0.01$ ‘**’, $p < 0.05$ ‘*’, and marginal at $p < 0.10$ ‘†’.

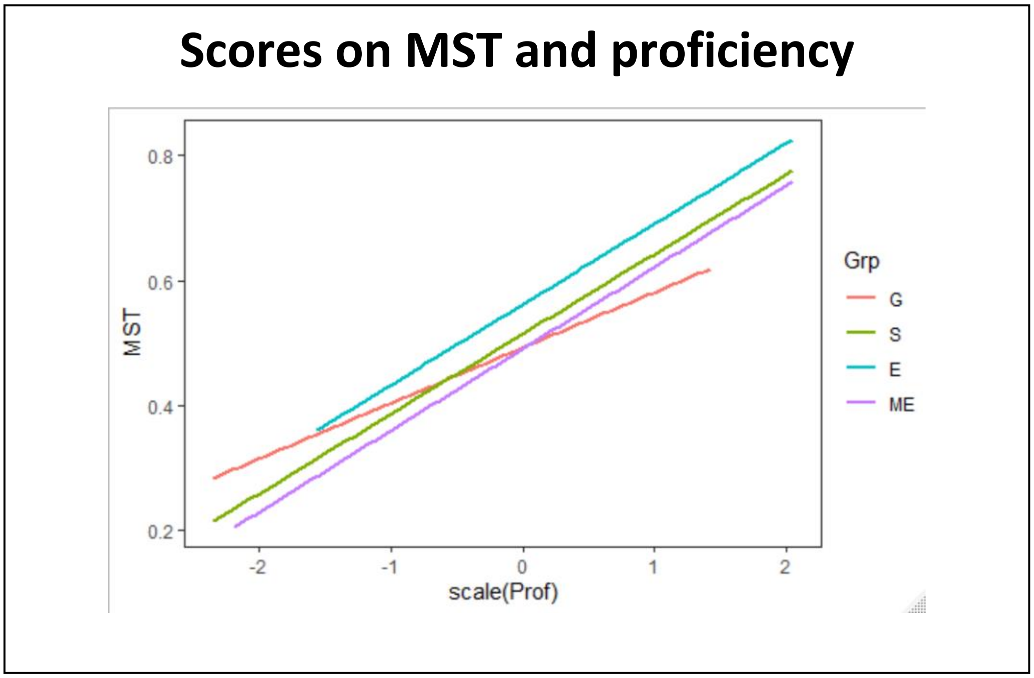


Figure 21. Profile plot of relationships between group scores on MST and proficiency.

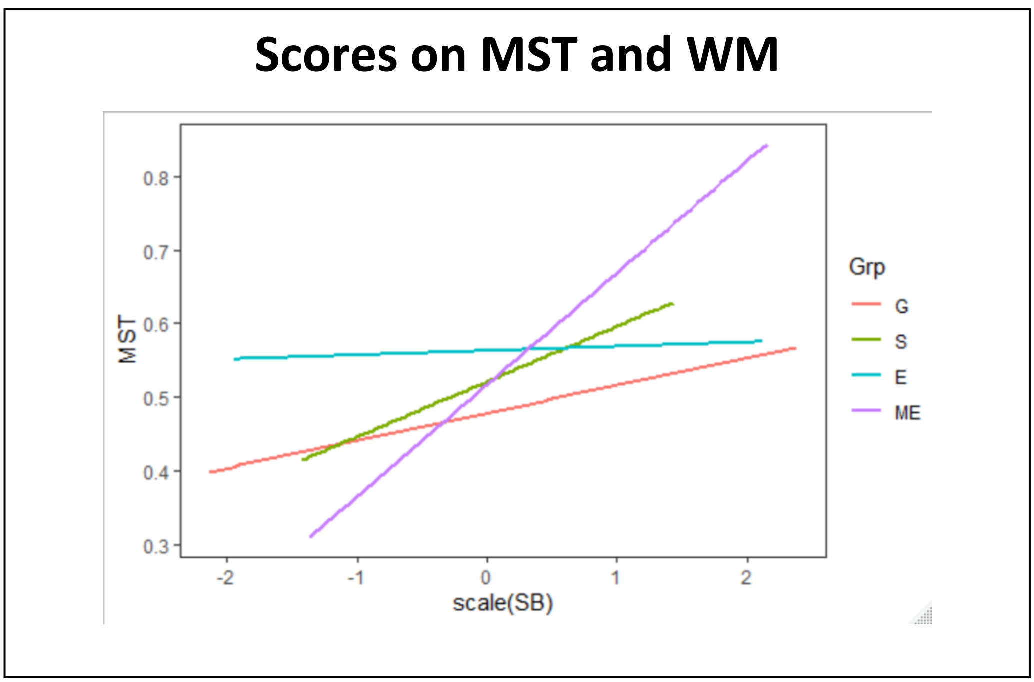


Figure 22. Profile plot of relationships between group scores on MST and WM.

5.6.3 Logistic regressions for main effects of input type on MDT

The final logistic MEM was performed with MDT binary scores as the response variable. The explanatory variables were input type, proficiency, and WM. The results showed the model that included the interaction term, Input typex Proficiency (AIC = 3912.5, BIC = 3979.9, -2 log likelihood = -1945.3, deviance = 3890.5, Chi-square = 68.23), to be the best fitting model. The model shown in Table 22 suggests that the elaborated input had the strongest effect on MDT scores. Additionally, the simplified input also showed a weak effect. The simplified group's MDT scores showed a significant interaction effect, and the elaborated group's scores showed a marginally significant interaction effect with proficiency. Furthermore, both moderator variables showed statistically significant effects on MDT (z = 2.66, p < 0.01; z = 2.80, p < 0.01).

The profile plot in Figure 23 shows the superior outcomes from the elaborated input group in comparison to the other groups, including the genuine input group. The plot in Figure 24 reveals that the elaborated group showed almost no effects of WM. There was no interaction effect between modified elaborated group and WM on MDT. These findings suggest that for definition recognition of TWs, while the elaborated input was best, the simplified input was better than the genuine and the modified elaborated input for participants at higher proficiency levels. Overall, for the meaning recognition tests (MST and MDT), the simplified and the modified elaborated input resulted in lower scores for participants at lower proficiency levels than the genuine input. These results are unexpected.

Table 22

Results of Logistic MEM for Input Type on MDT

*Best Fitting Model [MDT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered Prof + (1|Prsn) + (1|TW)]*

Fixed effects	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	0.36	0.21	1.74	0.081†
Simplified	0.36	0.19	1.89	0.059†
Elaborated	0.55	0.19	2.88	0.004 **
Modified Elaborated	0.18	0.19	0.97	0.331
Proficiency	0.33	0.13	2.66	0.008**
WM	0.19	0.07	2.80	0.005**
Simplified x Proficiency	0.41	0.18	2.30	0.021*
Elaborated x Proficiency	0.36	0.21	1.68	0.093†
Modified x Proficiency Elaborated	0.17	0.18	0.96	0.337

Random effects	Variance	SD
Intercept |Participant	0.30	0.55
Intercept |TW	0.79	0.89

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

Figure 23. Profile plot of relationships between group scores on MDT and proficiency.

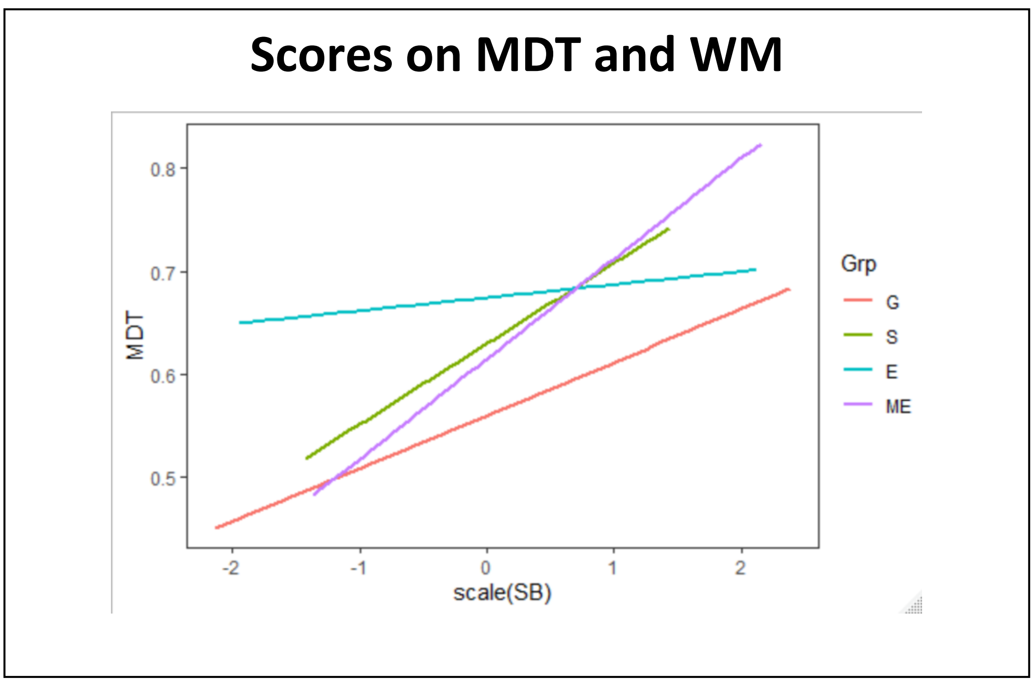


Figure 24. Profile plot of relationships between group scores on MDT and WM.

5.7 Debriefing results

After the vocabulary tests and the final WM task, debriefing sessions were held in which the participants were interviewed. The purpose of these sessions was to learn whether the participants had focused on content and whether they had noticed the TWs while listening to the talks. The results are shown in Table 23. Only four participants replied that they had focused on vocabulary, not content, while listening to the talks. Six participants replied that they had focused on both content and vocabulary. The rest of the participants reported focusing on the content of the talks. These responses indicate that the IVL condition was likely established, which means that, by and large, the results of the vocabulary post-tests reflect the by-product of another learning outcome during the primary activity.

As to noticing the TWs, 70% of the genuine version group answered that they had not noticed the TWs, whereas more than half of the participants in the input modification groups answered that they had: elaborated group, 64%; modified elaborated group, 57%, and simplified group, 54%. With regard to whether the participants' report that they had noticed the TWs was related to their scores on the vocabulary post-tests, statistically significant relationships were found with the FRT and MDT scores, but not with the MST scores (see Appendix M for correlations).

When the participants were asked why they had noticed the TWs, the most common reasons (n = 38) they gave were the TWs' unfamiliarity and/or the participants' difficulty in understanding the items (G: 7, S: 10, E: 9, ME: 12). Many responses implied that the difficulty originated from the specialization of the vocabulary items. According to 16 participants, known words were also noticed (G: 1, S: 3, E: 4, ME: 8). However, it was unclear whether the participants actually "knew" the words, for a number of reasons. First, during the debriefing sessions, those who claimed to know some of the words could not recall them. When pressed, the participants' responses showed that they often had misunderstood the meanings of the words or remembered the forms inaccurately. Also, by the end of the session, they had heard the TWs five times. Therefore, memory traces would be stronger during the debriefing. Only two participants responded that they knew about two thirds of the TWs but they did not know about one third of them. In the two elaborated groups, 12 participants responded that they had noticed the unknown words followed by the appositive cue *tsumari* 'i.e.' (E: 6, ME: 6). Very few participants responded clearly to follow-up questions regarding how they had noticed the TWs. Overall, their responses indicate that they were preoccupied with the primary activity of focusing on the content of the talks in order to answer the comprehension questions. Some remembered noting possibly important words that they had heard several times, and mentioned that they had tried to infer the meanings of some of these words in order to understand the content of the talks. Some participants reported that when they heard unknown specialized vocabulary items, they tried to imagine the kanji or think of English equivalents. The number of TWs that were known words to the participants was low overall, according to the participants' responses. Therefore, it appears that both unknown and known words could become salient during attempts to comprehend the input.

Table 23

Responses in Debriefing Sessions

Group	Focused on:			Noticed TWs:		
N = 106	Content	Vocabulary	Both	Noticed	Unnoticed	Unclear
Genuine	24	2	1	7	19	1
n = 27	(89)	(7)	(4)	(26)	(70)	(4)
Simplified	24	0	2	14	12	0
n = 26	(92)	(0)	(8)	(54)	(46)	(0)
Elaborated	23	1	1	16	9	0
n = 25	(92)	(4)	(4)	(64)	(36)	(0)
Modified Elaborated	25	1	2	16	12	0
n = 28	(89)	(4)	(7)	(57)	(43)	(0)

Note. Percentages (%) in each group appear inn parentheses.

5.8 Summary of findings

The study was designed to investigate the effects of input modification in the aural mode alone. It asked eight research questions. Table 24 summarizes the findings, and the following two subsections briefly review them, dealing first with the questions related to comprehension and then with those related to incidental vocabulary learning.

Table 24

Summary of Findings

Group	Results	Covariates		Interactions
N = 106		Prof	WM	
CQ Group	E > S > ME > G	***	**	
k = 36				
CQ Item	Inf > Rep > Syn	***	**	
k = 36				
Replication	ME† > S > E > G	***		
k = 12				
Synthesis	E > S > ME > G	***		S x WM**
k = 12				ME x WM†
Inference	E > S > G > ME	**		ME x Prof x WM†
k = 12				
FRT	E† > ME > G > S	***		ME x WM*
k = 32				
MST	E† > S > ME > G	**		ME x WM†
k = 32				S x Prof x WM†
MDT	E** > S† > ME > G	**	**	S x Prof*
k = 32				E x Prof†

Note. Significant at $p < 0.001$ '***', $p < 0.01$ '**', $p < 0.05$ '*', and marginal at $p < 0.10$ '†'.

5.8.1 Regarding comprehension

The first research question asked whether listening to modified versions of spoken input affects L2 listening comprehension, when L2 proficiency and WM are controlled. The outcomes of the comprehension questions showed that the participants who listened to the elaborated version scored higher than the participants who listened to the simplified, modified elaborated, or genuine versions, although the differences were not statistically significant. The participants who heard the other two modified versions also scored higher than those who heard the genuine version, but again, not at a statistically significant level.

The second research question asked whether specific types of input modification affect L2 listening comprehension differentially, when L2 proficiency and WM are controlled. Recall that the 36 comprehension questions were of three types: replication, synthesis, and inference. The results of a comparison of the scores on the different item types showed that the mean score of inference items was the highest and that of the synthesis items the lowest, but there was no statistically significant difference between the item types.

The results of separate analyses by item type showed that the hypothesis that the simplified group would perform better than the other groups on the replication items in the CQs was not borne out. Instead, the scores of the modified elaborated group were marginally significantly better than those of the other groups on these items; the simplified group did do better than the elaborated group, but the difference was not statistically significant. Regarding the CQ synthesis items, the elaborated group scored highest and the genuine group scored lowest. Both the simplified group's and the modified elaborated group's overall CQ scores showed interaction effects with the composite WM scores, which suggests that larger WM capacities were necessary when the participants needed to integrate the information provided by the shorter sentences of these two conditions. The interaction effect was stronger in the simplified group. The difference between the two input versions with shorter sentences might suggest that the modified elaborated group gained an advantage from hearing repeated and paraphrased words and sentences. The CQ inference item scores showed no statistically significant difference between the four groups, although there was a statistically significant three-way interaction effect of group and the two moderator variables (L2 proficiency and WM) for the modified elaborated group only. Yano et al. (1994) and Oh (2001) both found better overall performance on comprehension tests from simplified groups than from baseline (genuine) groups, at statistically significant levels. They also reported statistically significantly higher scores on replication and synthesis items from the simplified groups. In Yano et al. (1994), however, the elaborated group obtained significantly better scores on inference items. Oh (2001) included two different proficiency groups, and the low proficiency group performed better with the elaborated version than the simplified version in terms of overall CQ scores, replication items alone, and inference items alone, with statistically significant differences from the baseline group. Her results suggested that L2 speakers with lower proficiency might gain more from elaboration than speakers with higher proficiency. The results in the present study, however, suggest that elaboration can be beneficial for advanced L2 speakers as well.

The third research question asked whether L2 proficiency would moderate listening comprehension with the different types of input modification, when WM was controlled. Proficiency was included in all models pertaining to comprehensibility as a moderator variable. All the best-fit models showed proficiency to be a significantly strong predictor of listening comprehension and scores on all items. In the profile plots, proficiency was found to have played a role in higher scores in all input groups, showing clear linear relationships. This finding aligns with those reported in the prior literature (Chiang & Dunkel, 1992; Oh, 2001).

The fourth research question asked whether WM moderates listening comprehension with the different types of input modification, when L2 proficiency is controlled. This study is the first, to the researcher’s knowledge, to test the relationship between input modification and WM. The study hypothesized that the group that heard the elaborated versions of the listening passages would be most affected by WM, due to the syntactic complexity, larger number of sentences, and greater length of each passage. The composite WM scores were included in all models as another moderator variable. Findings from the best-fit models showed that WM specifically interacts with input modification and item types. First, WM was found to be a statistically significantly strong predictor of scores on CQs in both models that included effects of input modification type and item type. This finding indicates the role of WM to be as important as the role of proficiency. In models for each item type, however, WM was not significant with replication items. With the synthesis items, the role of WM was significantly important in the simplified group, and marginally important in the modified elaborated group. With the inference items, WM was not found to be important, except for a marginally significant three-way interaction with the modified elaborated group and proficiency. This finding is puzzling as it conflicts with the trend shown in the profile plots, in which

WM capacities appear to be more involved in scores on inference items than in scores on other item types, regardless of input modification type. Furthermore, the results shown in Figure 18 are mixed: The results for the replication and the synthesis items suggest that the groups that heard input versions with shorter sentences (i.e., simplified and modified elaborated) required higher WM capacities to obtain better scores on the CQs, while the other two groups (i.e., genuine and elaborated) did not. Crucially, the modified elaborated group showed a marginal three-way interaction effect with the moderator variables (WM and proficiency). These results show that the new input modification type (i.e., modified elaborated) employed by the present study required different cognitive processing than the other three input types, even though the information contained in the modified elaborated version was exactly the same as the information in the elaborated version.

5.8.2 Regarding incidental L2 vocabulary learning (L2 IVL)

The fifth research question asked whether input modification affects L2 IVL, when L2 proficiency and WM are controlled. On the form-recognition test (FRT), the elaborated group scored highest, showing a marginally statistically significant difference from the genuine group. Although the modified elaborated group scored the second highest, the significant interaction effect with WM indicates that participants with higher WM capacities performed better on the test. The simplified group scored lower than the genuine group, a finding that contrasts with results previously reported in the literature. Due to varied research designs in terms of types of input modification, input modality, types and frequency of TWs, and outcome measures, comparing results from the different studies is not easy. However, the L2 proficiency of the participants in the current study was higher than that of the participants in the other studies, and it appears possible that simplification has different effects on processing at different proficiency levels. The simplification might have caused information dilution, compared to the other versions. As a result, simplification may be less of an advantage at higher proficiency levels even for form recognition. The question of how simplification affects the processing of TW forms by L2 advanced speakers remains unanswered, and is worth investigating further in future studies.

The form-meaning sentence test (MST) scores showed a marginally statistically significant effect by group for the elaborated group, and a marginally statistically significant interaction effect with WM for the modified elaborated group. Thus, the participants with better WM performed better on the MST, as they also did on the FRT. For the simplified group's MST scores, there was a three-way interaction effect of group and the two moderator variables. This result could be interpreted to indicate that the simplified group's MST scores were more unpredictable due to the involvement of both proficiency and WM. In other words, even a highly proficient participant in this group could score low on the MST if the participant had low WM capacities. The MST was possibly the most cognitively demanding test in this study. The participants took it after they had heard all four listening passages. In the MST, they heard one TW and then four sentences with the TW; to make the correct choice, they had to immediately compare the contextual information in each sentence with what they had heard in the relevant listening passage, in order to select the sentence in which the TW was used with the same meaning. As their lexical knowledge of the TW was assumed to be fragmental, the sentence choices provided environmental cues for the participants to search for a trace of semantic memory of the TW. Therefore, the results of the modified elaborated group and the simplified group could indicate that shorter sentences provided a different quality of facilitative information for meaning recognition of the TWs than did the longer sentences heard by the elaborated and genuine groups. In processing unknown lexical items in the listening passages, the shorter sentences might have required additional WM effort to connect pieces of information.

The form-meaning definition test (MDT) scores showed two interaction effects with proficiency. For the simplified group, the interaction with proficiency was statistically significantly different, while for the elaborated group, the interaction effect with proficiency was marginally significant. These results differed from the results for the other post-tests. On this test alone, WM was a significant predictor of test score, which suggests that WM capacities were required to select the most appropriate definitions of the TWs. While previous literature has reported that groups in a simplified input condition perform better on form recognition tests than those in a genuine input condition, the finding in this study indicates that the simplified group participants performed better on the meaning recognition test than on the form recognition test when their proficiency was higher.

In summary, all input modification groups scored higher than the genuine input group on the unannounced vocabulary post-tests except for the FRT. The prediction that the modified elaborated version would lead to better performance than the other input modification types was not supported. The simplified group scored higher than the modified elaborated group except on the FRT. Neither modified elaborated input nor simplified input showed statistically significant effects on the groups' TW post-test scores.

The sixth research question asked whether specific types of input modification would affect L2 IVL differentially, when L2 proficiency and WM were controlled. The study predicted that the modified elaborated group would perform significantly better than the other groups on the meaning recognition tests. However, the modified elaborated group performed better at a statistically significant level on the FRT only when participants had higher WM capacities. The modified elaborated group's scores also showed a weak interaction effect with WM on the MST, but not on the MDT.

The seventh research question asked whether L2 proficiency would moderate L2 IVL with the different types of input modification, when WM was controlled. The effect of proficiency was statistically significant across the post-tests. Interaction effects of proficiency with simplified input and with elaborated input were found on the MDT, which suggests that the L2 definition test scores were associated with proficiency in these groups.

The final research question asked whether WM would moderate L2 IVL with the different types of input modification, when L2 proficiency was controlled. Among the three post-tests, scores on the definition test (MDT) alone showed a significant relationship with WM for all input groups. In other words, when participants had to select the most appropriate definitions for the TWs, WM capacities were found to be a significant predictor of their scores. Scores on the FRT and MST showed a relationship with WM for the modified elaborated group, and scores on the MST showed a marginally significant relationship with WM and proficiency for the simplified group. These results suggest that the modified elaborated and the simplified groups required more WM to process the input than other groups, as shown by the groups' scores on the CQs, as well.

Chapter 6 Discussion

The present study investigated the effects of auditory input on listening comprehension and incidental L2 vocabulary learning (L2 IVL) by advanced L2 Japanese speakers. The study delves into some under-researched areas in the IVL literature. Specifically, it ends a hiatus in research on input modification for listening (in contrast to reading) comprehension. The study is also unique in investigating IVL from spoken input alone. The existing literature suggests that L2 IVL gains from listening are lower than those from reading. This study investigated the possibility of greater IVL gains from three different types of input modification, all with three exposures to each target word (TW). As a novel contribution, the study examined a new type of input elaboration, a modified elaborated version, along with an elaborated version and a simplified version, all three compared to an unmodified baseline version. Previous studies have shown the facilitative effects on comprehension of modified input (e.g., simplified, elaborated) compared to those of unmodified input, that is, genuine input. The modified elaborated condition in this study contained rich contextual information with shorter sentences, thus offering, potentially, a middle ground between elaborated and simplified versions. Furthermore, WM and L2 proficiency were included as moderator variables.

6.1 Primary findings

6.1.1 Robust effects of elaborated auditory input and learner-related variables

The most important finding of this study was that elaborated input had robust effects on test scores. Although the effects were not consistent in terms of statistically significant differences from the effects of the genuine input, both the mean test scores and the profile plots showed consistently better outcomes from the elaborated than the other input types, except for scores on the CQ replication items. The elaborated input used was the full version, including both lexical elaboration, such as synonyms and definitions, and structural elaboration, such as added regularity, paraphrasing, and repetitions, at the sentence level. The lexical elaboration was added for the purpose of clarifying the meaning of lexical items, whereas the structural elaboration was designed to clarify the content and the semantic flow of the talks. It was more syntactically complex than the simplified or modified elaborated input, as the results of the Mann-Whitney U Test showed. The findings, however, favored full elaboration.

The purposes of asking comprehension questions were: (a) to replicate the study by Yano et al. (1994) that found relationships between input modification and comprehensibility in reading, and (b) to establish an IVL condition by keeping participants' focus on the content of the input.

The results of the replication part of this study did not align with those reported by Yano et al. (1994). Their results support both simplification and elaboration as beneficial for reading comprehension by intermediate L2 speakers. In contrast, the simplified group in the current study did not perform better than the elaborated group; nor did elaborated input lead to statistically significant effects on CQ inference item scores. This difference in results may lie in the modality difference; this study's finding of a significant correlation between CQ responses and composite WM suggests that, although elaborated input mitigated the level of the demand, the auditory input tasks were cognitively quite demanding. In short, there is no easy comparison between the findings from auditory input in the present study and those from written input in Yano et al. (2004), which might be based on different cognitive processes and load.

The different results in this study and previous studies also may be due to differences in participants' L2 proficiency. This study's participants were recruited from local Chinese communities in Tokyo, while most participants in previous studies have been ESL/EFL students at universities and high schools (Chung, 1995; Oh, 2001; Urano, 2000; Yano et al., 1994). Thus, the participants in previous studies have been relatively homogeneous: students in EFL/ESL contexts with limited exposure to the target language (English). In contrast, participants in the current study had been exposed to the target language (Japanese) in their daily lives for at least six months. As a result, there was a substantial difference in L2 proficiency between these participants and those in prior studies, and this difference might be reflected in the current study's findings of effects of elaborated input and the unpredicted weaker effects of simplified input. Furthermore, the current results supported the value of elaboration for high proficiency speakers, even those who were likely to be exposed to genuine input in their daily lives.

For IVL, the elaborated input group performed better than the other groups, but all input modification groups achieved higher scores than the genuine input group on all the post-tests, except for the simplified group on the FRT. In the recent IVL literature, input modification with auditory input has not been used as a treatment. The study employed a full range of elaboration techniques—adding definitions, synonyms, and/or exemplification as lexical elaboration, and adding repetitions, paraphrasing, restoration of canonical word order and/or dropped pronouns, and summary statements as structural elaboration—to provide rich information about the TWs. The findings suggest that the additional information was beneficial for IVL, and did not overly tax WM, upon hearing unfamiliar lexical items. This study's treatment consisted of listening to highly specialized content for approximately 60 minutes, and while the comprehension questions were presented in written form, all the multiple-choice items on both the comprehension questions and the vocabulary post-tests were presented only auditorily. Thus, the findings of the study are encouraging for the potential of auditory interventions for advanced L2 speakers, given that the previous literature suggested low IVL gains from listening alone.

Typographic input enhancement is still generally considered a stimulus for implicit learning in the L2 literature, although recent results suggest its effects are explicit, not implicit (Borro, 2020; Pellicer-Sánchez & Boers, 2019). Its effects in the previous studies appeared to be largely limited to form recognition. In contrast, the finding of the elaborated input group showed learning of both the meanings and forms of the TWs from IVL during a short listening treatment.

6.1.2 Proficiency and WM

Due to the use of only auditory input, it was expected that both moderator variables, proficiency and working memory, would play important roles in the outcomes. They were found to be significant predictors of listening comprehension and IVL.

Previous IVL research has primarily focused on the effects of proficiency (Chaudron, 1983; Long, 1985). While elaborated input is widely considered more beneficial for L2 speakers whose proficiency level is lower, the results reported in the literature do not necessarily contribute to a clear picture of the relationships between input modification and proficiency, due to the variety of research designs and L2 speaker populations. The proficiency of the participants in the study was advanced. The results from logistic MEM showed significant effects of proficiency on all tests. Among the tests, only the MDT results showed significant interaction effects of proficiency with both simplified and elaborated input. The findings suggest that the advanced L2 speakers also favored elaborated over genuine and simplified input.

Parker and Chaudron (1987) reviewed early input modification studies and summarized the robust results for elaboration in contrast to the mixed results for simplification. They assumed that elaboration would be more beneficial for L2 learners at lower proficiency levels, citing studies in which the most advanced learners performed equally well on both genuine and elaborated versions. Furthermore, based on findings from a reading comprehension study, they suggested elaborated text could be a more natural alternative to genuine text than simplified text, by pointing to the fact that the syntactic complexity of elaborated versions did not interfere with learners' comprehension. In other words, Parker and Chaudron did not view elaboration as taxing cognitive processing capacity, possibly including WM, which aligns with the findings of the current study regarding the relationship between elaboration and WM.

Based on the findings, elaborated input could consistently provide L2 learners with better contextual information that enhances comprehension and IVL regardless of proficiency. It appears to be widely believed that simplified input should be easier to comprehend due to the short sentence length and use of higher frequency words, which may be why simplified input is so often used for L2 learners with low proficiency. Yet the mixed results from simplified input in the previous literature and this study suggest that elaborated input could provide higher quality input that would be more beneficial to L2 learners at a wider range of proficiencies.

As to WM, Wen and Li (2019) pointed out the lack of studies investigating relationships with IVL. They reported that L2 vocabulary learning studies so far have investigated the role of WM based on experiments with rote memorization tasks. Malone (2018) also mentioned that L2 vocabulary learning studies with WM tasks generally adopt explicit rather than implicit learning conditions.

Therefore, this study contributes to filling a gap in the literature by adopting incidental learning conditions with listening comprehension tasks and assessing WM. The study employed two non-verbal complex WM tasks. As shown in Appendix H, scores on Shapebuilder showed significant correlations with scores on CQ, all vocabulary tests, and the proficiency test. In contrast, OSPAN scores showed significant correlations with CQ and Shapebuilder scores. Linck et al. (2014) highlighted the importance of including WM, and investigating its relationship with proficiency, in SLA research. The findings of this study support the importance of WM, although the results related to WM were mixed, with slight differences depending on the task.

The scores on CQs were significantly correlated with both WM tasks, whereas the scores on vocabulary tests were significantly correlated with Shapebuilder alone. For the CQs, participants listened to a part of each talk and then saw a question on the computer screen; they selected possible answers from four choices provided in the auditory modality only. Each talk was divided into three sections and participants' comprehension was tested incrementally. This format was selected to ensure that participants' attention would be drawn to the content of the talks without taxing their WM.

The findings pertaining to the relationship between WM and input modification were surprising. The study predicted that elaborated input would require more WM capacities due to its greater syntactic complexity and longer sentences. The modified elaborated input was expected to compensate for the possible cognitive overload of the elaborated input by breaking up the information into short sentences. Countering these predictions, the profile plots showed little WM influence on the elaborated group’s performance, and more WM influence on that of the modified elaborated group. Furthermore, both CQ item type and input modification type were found to interact with WM in different ways. The findings pertaining to comprehension question item-types showed clear differences in cognitive processing. Specifically, inference items required WM resources regardless of the input modification type. This suggests that responding to these items is more cognitively demanding because they require participants to develop judgments on the basis of knowledge they would have just integrated from the talks. Conversely, WM capacities did not influence the synthesis item scores of the elaborated and genuine input groups, nor the replication item scores of the elaborated input group. WM capacities did, however, affect the replication item scores of the modified elaborated input group, who performed better on replication items than any of the other groups. These findings contribute to a clearer picture of the roles of these moderator variables in advanced L2 speakers’ cognitive processes.

6.2 Better elaboration: Why did modified elaboration not work as expected?

The results from the modified elaborated input were not as predicted. This new type of modification included exactly the same information (e.g., vocabulary items) and the same level of syntactic complexity as the elaborated input. The only difference was sentence length, which was shorter in the modified elaborated version. The findings revealed that the processing of modified elaborated input was more comparable to the processing of simplified input than to the processing of elaborated input, in terms of WM requirements. The effect was found especially in responses to CQ replication items. In contrast, simplified input did not show significant effects by item type. When larger WM capacities were available, modified elaborated input was found to be the most effective on FRT and marginally more effective on MST scores. Elaborated input was more effective on both tests. Thus, these two input types show comparable trends. The findings regarding MDT were puzzling because they showed opposite results from those obtained for elaborated input and modified elaborated input. Only for this test, simplified input was the most effective when proficiency was higher.

L2 sentence processing research has shown that L2 speakers require more WM capacities when processing long, complex, or ambiguous sentences (Jiang, 2018). Complexity here includes syntactic complexity. Given the role of WM, the amount of information in the input also affects sentence processing, and the balance between sentence length and the density of information in the listening passages might help explain the current results for modified elaborated input. Chaudron and Richards (1986) reported that ESL students with English proficiency at approximately the university entry level comprehended a lecture that included macro discourse markers better than lectures without markers or with other markers (micro, both micro and macro). They suggested that the students may have paid attention to macro markers as a way of supporting their comprehension, based on the overall organization of the lecture. This study also used talks that contained dense information with specialized vocabulary. Due to the nature and length of these talks, the participants might have paid more attention to the overall organization and the cohesion of information. A focus on the overall flow might lead to a need for more WM capacities to process the meaning of the short but elaborated sentences of the modified elaborated version.

Koda (2005) summarized a major controversy regarding text complexity and comprehension difficulty. The complexity of a text is determined by linguistic features, such as vocabulary and sentence structures. Although researchers often use the length of words and sentences to analyze text complexity, Koda suggested that word and sentence length might not be direct causes of comprehension difficulty. She argued that syntactic complexity provides a clearer indication of semantic relatedness among elements in a sentence, such as clauses and nouns. Thus, it facilitates better processing and comprehension. Specifically, Koda pointed out that subordination assists comprehension by adding information pertaining to connections between clauses. Without subordination, the messages behind sentences (clauses with subordination) could remain opaque, because clear links between sentences are missing. Syntactically complex sentences can render logical connections, such as causal and contradictory relationships, more transparent, and assist semantic integration in processing within a discourse. Koda concluded that it is not reasonable to assume an automatic link between features of a passage, such as syntactic complexity and sentence length, and comprehension difficulty, regardless of input modality. She suggested that whether such features affect the semantic and logical cohesiveness of a passage is influenced by the speaker's or writer's underlying intention.

As elaboration studies began investigating how NSs adjust their speech for NNSs (e.g., Chaudron, 1985a, 1985b; Long, 1983a, 1983b), some of their results could be interpreted as indicating that the long sentences in elaborated input have a side effect: The very fact that a message is embedded in a larger structure makes that message clearer through the use of the additional words, phrases, and subordinating structures that the longer sentences require. In this study, several participants who listened to the modified elaborated version reported difficulty understanding the passages' logical cohesion. Their remarks in the debriefing interviews suggest that they experienced all the additional cues as disrupting the flow of the talk when these cues were part of short sentences. In other words, in short sentences, lexical elaboration becomes too salient, hindering the processing of underlying information. Additionally, the content of the talks was not easy even for the advanced L2 speakers to understand immediately. The information in the short sentences was felt to be too dense to understand, while the relationship between the short sentences became opaque, despite the added inter-sentential connectors in the ME version. Davey (1988) argued that sentence length and passage coherence are closely associated, because longer sentences are likely to include explicit connectives, such as *because* and *although*, and thus texts with longer sentences will be clearer and more coherent than those with shorter sentences.

Long (2019) also pointed out how choppy sentences could tax cognitive processing and result in the incomprehensibility of meaningful links in a passage. The ratio of his example sentences in the four versions was five sentences in the simplified, two in the elaborated, and three in the modified elaborated for just one in genuine input.

In this study, the number of sentences in the modified versions was approximately comparable to that in the simplified versions.

Derwing (1996) explored four types of input modification to understand which might be more beneficial for NNSs and which might instead disrupt comprehension. The four conditions were: (a) no elaboration added; (b) paraphrasing with a clear, explicit marker, such as *in other words*; (c) paraphrasing without a marker or with an ambiguous one; and (d) irrelevant elaboration added. Across three experiments, participants showed comprehension difficulty as a result of the irrelevant elaboration type. Therefore, while Derwing's results supported findings from previous studies regarding the positive role of elaboration in L2 comprehension, she argued that both quality and quantity of elaboration should be further investigated for L2 comprehension and learning. She emphasized the need to distinguish psycholinguistically sound elaboration and irrelevant elaboration that could cause cognitive overload.

In summary, the current results appear to have produced two main findings: (a) elaborated input is effective in improving both comprehension and IVL; and (b) modified elaborated input requires more WM capacities, due to the short length of the sentences. In contrast, elaborated input was found not to be influenced by WM capacities. The findings support elaborated input as a psycholinguistically sound intervention that is unobtrusive yet effective. Although the effects of modified elaboration should be further explored with longer sentences, the current results for elaborated input in a listening alone treatment are robust, supporting previous research findings. To understand how best to modify elaborated input for better intervention, however, will require further exploration and consideration of a wider range of variables that are likely to be involved, such as text-related and L2 speaker-related variables.

6.3 Lexical knowledge and quality of input: Enhanced IVL and semi-IVL conditions

Remarks made by the participants in the debriefing interviews provide a peek into how they attempted to connect unknown vocabulary items with the content of the talks. Just as the revised hierarchical model (RHM) hypothesizes, this study's participants used conceptual links while listening to L2 spoken input. Some participants described how they searched for possible kanji for unknown TWs based on the phonology. Because Japanese includes many homophones, it was not easy for them to accurately match phonological cues and kanji compound words, and selecting the wrong kanji would lead to incorrect answers on both comprehension questions and vocabulary tests. Due to their advanced proficiency level, they could recall a fairly large number of kanji, which meant the probability of selecting the wrong kanji combinations was not low. Furthermore, based on information gathered in the debriefing interviews, it appears that once they selected a kanji, it was difficult for participants to review and revise the association. Having an incorrect kanji in mind easily leads to selecting an incorrect kanji compound word, which could have a large impact on comprehension, due to the amount of information in the compound word. L1-Chinese learners of Japanese are usually considered to have an advantage, due to their knowledge of Chinese characters, from which kanji derive. The participants' comments in the debriefing interviews, however, suggest that Japanese homophones frequently misled the participants as they strove to understand the unknown words.

Whether the misleading homophones negatively affected their listening comprehension can be treated as a separate issue. Participants likely encountered ambiguity of homophones when they listened to each TW on the vocabulary tests. In contrast, in the input, ambiguity was avoided due to the context and additional information available in the talks. Furthermore, lexical elaboration was helpful to grasp the meaning of low-frequency words, including TWs. Therefore, it is assumed that participants experienced less trouble with the issue of homophones in the talks.

According to data from the language background questionnaire, many participants were fluent in English, as well. Some reported that they regularly watched American TV dramas and movies in English. During the debriefing interviews, some mentioned searching for equivalent English words when they heard unknown English loanwords in the talks. Inferring the original meanings of English loanwords in Japanese, however, is known to be difficult (Daulton, 2008). Long and Ross (2009) posited a possible input modification study with L2 Japanese and pointed out the issue of text modification with ideographic writing systems. In fact, the Japanese writing systems are complicated, consisting of both syllabaries and ideographs. This issue was one of the reasons this study employed a listening comprehension task, excluding written texts as much as possible. Although the participants' subjective reports in the debriefing interviews are limited as a means to learn about their actual cognitive processes, they appear to reflect at least part of the participants' deductive processes pertaining to the Japanese writing systems.

When searching for the meaning of an unknown lexical item, participants attempted to use information from all available lexicons: not only their L1, but also their L2 Japanese, and possibly their L3 English. Simultaneously, they were linking concepts of which they already had knowledge to understand the meaning of the sentences. The contextual information in the talks must have provided useful cues in the process. Both lexical and structural elaboration are assumed to provide higher quality information. One participant, whose Japanese was very advanced, explained that he could remember an entire sentence containing an unknown word. In this particular case, the contextual information around the unknown word could support his word memory.

On the other hand, some participants claimed that they could remember new words if they knew the meaning, but they had not noticed or remembered the TWs in this study because they did not know their meanings. As mentioned in the previous chapter, however, the meaning of *to know* appeared to vary, depending on the words in question and the participants. Only two participants were confident enough to mention specific approximations. Others were vague in describing the state of their lexical knowledge, saying, for instance, that they had “heard it before,” but they did not claim to “know” more than a few of the TWs they had heard in the listening passages. Most participants described both the talks and the tests as difficult. As the debriefing interviews revealed, the TWs, which were specialized and low-frequency items, were unfamiliar to participants, as intended. Additionally, the ambiguity of “knowing a word,” which became clear in the interviews, suggests the difficulty of assessing whether a learner retains traces or weak mental representations of words encountered incidentally.

This fragmentary lexical knowledge can be explained by Jiang’s (2002, 2004) framework. His model assumes incomplete novel lexical entries from limited exposure, and claims that L1 translation plays a role in completing a lexical entry. However, in the case of advanced learners, the process may differ. As the RHM hypothesizes, an incomplete lexical entry may be filled with meaning from the L2–concept link. In this case, the semantic information may not be language-dependent. Additionally, how a novel lexical entry emerges to begin with is still a question. With contextual support, the initial entry may emerge at the first exposure in the case of an advanced L2 learner.

Semi-incidental learning conditions and enhanced IVL

Hulstijn (2003) argued for the need to maintain a clear distinction between incidental learning and implicit learning, but also acknowledged the difficulty of drawing that line. Implying that definitions of these types of learning could differ depending on the focus of research, he nevertheless pointed out a fundamental concept of incidental learning in contrast to intentional learning: "the involvement of attention is not deliberately geared toward an articulated learning goal in the case of incidental learning" (p. 361).

Pellicer-Sánchez and Boers (2019) acknowledged the difficulty of operationalizing the distinction between incidental and intentional learning. They argued that, in experimental research, the incidental learning conditions in which L2 speakers were led away from intentional learning could be stipulated. In an actual learning situation, however, they considered both types of learning to support and facilitate L2 development. Given the many studies conducted to investigate incidental and intentional learning for grammar, vocabulary, and formulaic language, and the variety of their research designs, Pellicer-Sánchez and Boers proposed a new category, which they called *semi-incidental learning* conditions: Learners are led to become involved in a text to comprehend it, while the salience of target lexical items is raised to heighten learners' awareness of them, using a variety of devices, such as typographic enhancement.

Long (2019) argued that implicit knowledge, of which learners themselves are not aware, is a potential outcome of incidental, but not of intentional, learning. He strongly argued for the importance of finding viable ways of facilitating incidental L2 learning, especially for vocabulary and collocations, to leverage L2 learners' proficiency toward the advanced level, given the time constraints in regular classrooms, which typically must cover textbook-based lessons with explicit instruction. Long (2017) explained the subtle differences between conscious noticing and subconscious (attentional) detection under incidental learning conditions, referring to Tomlin and Villa (1994). If implicit knowledge is the goal, the priority must be to facilitate subconscious detection. Hence, the solution proposed by Long (2017, 2019) is *enhanced incidental learning*, which describes unobtrusive interventions designed to increase learners' unconscious detection. Such a learning condition is more psycholinguistically appropriate to establish mental representations in learners' long-term memory compared to other consciousness-raising interventions employed in incidental L2 learning conditions.

The input modification used in this study was different from input enhancement (Pellicer-Sánchez & Boers, 2019; Sharwood Smith, 1993). Participants demonstrated vocabulary gains on both form- and meaning-recognition tests. This outcome contrasts with reports from input enhancement studies in the literature (see, e.g., Borro, 2020), which show input enhancement to be effective for learning forms, but not meanings, and for developing explicit, rather than implicit, knowledge.

With regard to unconscious detection, however, the results of the debriefing interviews might suggest that participants' consciousness might have been raised for TWs. Because a majority (90%) of the participants reported having focused on the content of the talks instead of vocabulary, it is safe to say that an incidental learning condition was established. Conversely, more than half of the participants who listened to modified input said they had noticed the TWs, as did a few in the genuine group. The most informative data from the interviews revealed a proportional difference among the groups in participants who noticed TWs. In the genuine group, only 26% of participants noticed the TWs. In contrast, in all the input modification groups, at least half of the participants noticed them, with the highest proportion in the elaborated group. Although unfamiliarity with and/or difficulty in understanding the TWs might have raised their salience, the low proportion of participants who noticed them in the genuine group suggests that this scenario is unlikely. Hence, the input intervention might not have been as unobtrusive as expected.

6.4 Pedagogical implications

As one of the major current research topics in SLA, IVL is the focus of a growing body of literature (De Vos et al., 2018; Uchihara et al., 2019), and there is a consensus that intentional learning and explicit instruction are limited in teaching the great amount of vocabulary L2 learners need (Hulstijn, 2003). Moreover, research shows that vocabulary knowledge is important as a medium for learners to develop grammatical knowledge (Martin & Ellis, 2012; Vafaee, 2016). The quality and quantity of input appear to be crucial for increasing incidental learning of vocabulary (Laufer & Hulstijn, 2001).

Based on theories and research findings in SLA and related fields, Long (2015) proposed 10 methodological principles in Task-Based Language Teaching (TBLT) that are versatile enough to be applied in any language class. One of them is to elaborate input, which, according to Long, "is psycholinguistically more appropriate than either genuine or simplified input" (p. 306). He also emphasized learners' need for rich input to develop functional L2 abilities. The current study examined this methodological principle by investigating the effects of elaborated input not only on comprehension, but also on IVL. The findings of this study overall support TBLT by aligning with those in the literature, and further revealed that advanced L2 speakers with input elaboration performed better than those with genuine input. These findings, as well as other information gathered in the study, could be invaluable in designing L2 course syllabi and materials.

More specific pedagogical implications can be proposed on the basis of the findings. First, in contrast to traditional L2 teaching materials in which either simplified or so-called "authentic" materials are primarily used, the findings showed that elaborated versions are better than those materials. Simplified input, which is widely believed to be the most comprehensive input, shows mixed results regarding its effectiveness in the prior literature and this study. It may increase comprehensibility through less syntactic complexity. However, the findings suggest that short sentences increase cognitive load. Thus, simplification may not be as facilitative as educators expect. Rather, the findings suggest that input elaboration should be more widely employed to provide better quality materials in L2 language teaching and learning. Many instructors, however, do not know how to create such materials. Therefore, a variety of elaboration techniques, including those at the lexical level with explicit signaling, paraphrasing, and repetition, should be clearly introduced to L2 instructors and material writers. The introduction should be thorough and practical for L2 classrooms, and, as Derwing (1996) suggested, the quantity and quality of elaboration should be carefully calibrated for target L2 learners. To enhance IVL, it is better to introduce the idea of selecting TWs according to available word-frequency lists rather than counting on instructors' intuition and experiences. Frequency band and number of occurrences of TWs should also be determined to create appropriate materials and tasks according to learners' L2 proficiency.

Consequently, it is important to emphasize the value of providing students with opportunities for learning the L2 incidentally through unobtrusive interventions. Due to time constraints and the need to cover designated textbooks and meet goals stipulated by schools and L2 programs, explicit instruction is conventionally considered the most suitable approach for the teaching and learning environments of regular L2 classrooms. Although much research encourages the inclusion of additional activities, such as extensive reading/listening and their online formats, in a curriculum or as extra-curricular activities, regular classes appear to lack room for incidental learning opportunities. However, if the focus of L2 classrooms shifts from teaching and learning how to use L2 forms to using the L2 for communication, L2 learners should encounter more opportunities for incidental learning. In such classes, whether meaningful information is successfully understood and conveyed among students using the L2 as a tool is crucial. In their exchanges in the L2, their attention is drawn to task completion, rather than linguistic forms.

Input elaboration should play an important role in such meaning-focused classes. The more detailed information students have, the more probable their successful completion of a task, which leads to a stronger sense of accomplishment, more satisfaction, deeper understanding, and thus, stronger motivation for L2 learning. Most importantly, this approach should enhance L2 development. As mentioned earlier, the use of simplified texts is pervasive where genuine materials are not used. Yet, as we and others (see, e.g., Al-Thowaini, 2018) have argued, simplified language risks not only providing poor quality input, but also interfering with students' comprehension of content.

One third of the participants were professionals in both private and public sectors. The rest were students planning to work in Japan. Due to Japan's aging society, more and more professionals and workers are expected to be L2 speakers of Japanese. Mass media, such as newspapers and TV, are already increasing their use of easier formats with less complicated vocabulary and sentences. There is growing public discussion of the need to create texts in "easy Japanese" (*yasashii nihongo*) and to improve the readability of Japanese, as matters pertaining to human rights. All these trends have led to the simplifying of Japanese input in the mass media and announcements from municipal governments in Japan (Iori, Iwata, & Mori, 2011; Lee, 2016; Tanaka & Mino, 2010). The focus, however, is on making information easier to comprehend, and not on facilitating second language acquisition and development (see Long, 2015, p. 250, for a comparable argument). In fact, the poor input in publicly available texts may result in a lack of good quality input for Japanese L2 learners. Therefore, it is hoped that the findings of this study add another layer of consideration regarding SLA in Japanese and make a contribution to this discussion.

Positive effects of elaboration on reading comprehension have been supported in the literature. The findings of the current study suggested that elaboration is also effective in the auditory mode alone. Furthermore, elaborated input is beneficial regardless of WM capacities. This finding also sheds light on text analysis tools, such as readability judgment tools. The tools analyze linguistic features, such as syntactic complexity, type and token of lexical items, and sentence length. However, they do not analyze semantic information and how it is conveyed in a text. Based on the results of input elaboration studies, including this one, readability may not be correctly assessed on the basis of linguistic features alone. With regard to receptive skills, the findings suggest that replication, synthesis, and inference comprehension questions draw on different cognitive processes. Therefore, all three types of questions are useful for assessing students' comprehension of reading and listening passages.

Chapter 7 Conclusions and Directions for Future Research

The current study revisited effects of input modification on comprehension with spoken input in Japanese, while also investigating IVL. The research design included both proficiency and WM as moderator variables, and data collected from 106 advanced L2-Japanese speakers were analyzed using mixed-effects modeling. It is the first to compare effects of elaborated and genuine input with advanced JSL speakers. It also examined effects of a new type of elaborated input, referred to as modified elaborated. The new type was predicted to be the most effective input modification for both listening comprehension and IVL, regardless of participants' proficiency and WM capacities. Modified elaborated versions of listening passages were created based on elaborated versions. Only sentence length was changed, by segmenting longer sentences in the elaborated version into shorter sentences comparable to those of the simplified version. Lexical and structural elaboration in elaborated versions were preserved. Flow of the talks was maintained by adding inter-sentential connectors.

The findings revealed that the elaborated input was effective for both comprehension and IVL. While results from the elaborated input condition did not consistently show statistically significant advantages, the highest scores of the elaborated input group on all tests but one are evidence supporting the provision of rich input. Although the participants were familiar with listening to genuine spoken input, having lived in Japan for half a year or longer, the results showed that the elaborated input group performed better overall than the genuine and simplified input groups. These results demonstrate that input elaboration was beneficial even for very advanced L2 speakers, which was newly validated in the present study.

The three types of input modification showed variance in their results; some aligned with previous findings, whereas others did not. Despite the efforts to preserve all of the elements of the elaborated input in the modified elaborated versions, the results from the elaborated input and modified elaborated input groups differed. Unlike the elaborated input group, some results from the modified elaborated group were comparable to those of the simplified group. The simplified group showed mixed results relative to the genuine and elaborated input groups. This finding with respect to simplified input aligns with findings in previous studies.

The findings pertaining to L2 proficiency were consistent across all tests. Proficiency showed significantly positive effects on both comprehension and IVL. The participants were more proficient than those in previous studies. Significant and marginally significant interaction effects with proficiency were found only on MDT with simplified and elaborated input. Overall, proficiency was found to be the strongest predictor of all results.

With regard to WM, the findings suggested different cognitive processes according to input modification type and comprehension question item type. Input modifications that involved shorter sentences required more WM capacities, while the elaborated and genuine versions, which had longer sentences, did not require as much WM. Especially significant interaction effects with WM were found on FRT with the modified elaborated input, and on synthesis items with the simplified input. These findings were unexpected and appear to contradict a general belief that longer sentences require more WM. Therefore, further investigations are desirable. In particular, the effects of a revised version of modified elaborated input with longer sentences should be investigated further, in order to explore a medium potentially more conducive for enhanced IVL.

Limitations of the study

The participants were a diverse group of advanced L2 speakers resident in Japan. In contrast with previous studies whose participants were all students, in this study, one third of the participants were not students. A variety of resources and word of mouth were employed to recruit these diverse participants. However, they were not a random sample; rather, the method of recruitment was convenience sampling, and compensation was offered. Therefore, the generalizability of the results is limited.

Lack of strong statistical evidence for effects of input modification on both CQs and IVL is another limitation of the study. Treatment sessions were short and intense, with four academic mini-talks. This may explain the weak statistical evidence for effects of input modification. A longitudinal study would provide stronger evidence regarding whether the different treatments could have statistically significantly different effects.

Some limitations pertaining to the tests should be mentioned. First, pre-tests were not employed to examine participants' pre-existing knowledge of TWs or their Japanese vocabulary size. Primary reasons for not doing so were (a) to avoid exposing participants to TWs, and (b) lack of an appropriate vocabulary size test. Instead, low-frequency words in specialized fields were selected for the research (Hulstijn, 2003). Possible candidates for TWs were chosen directly from texts written by experts. As described in Section 4.2.1, the talks used highly specialized vocabulary items. To assess incidentally learned vocabulary items, 32 target nouns were carefully selected based on their origins; kanji compound words that originate from Chinese were avoided as cognates, even when their Japanese pronunciation was dissimilar to the equivalent in Chinese. Some English loanwords were included as target words due to the difficulty of deciphering them in Japanese on the basis of English knowledge only. The target word list was revised based on the results from two pilot sessions with highly advanced L1-Mandarin Chinese L2-Japanese speakers. Their comments during the debriefing sessions were informative to create the final list. In addition, in the process of selecting TWs, Japanese word-frequency lists were consulted to finalize the TWs. As a result, most participants did not know them, as validated during the debriefing sessions (see Section 5.7 for debriefing results). As discussed above, there appear to be many incomplete lexical entries in L2 speakers' lexicons. It is assumed that the participants, who were advanced L2 speakers, must have had a large number of such incomplete lexical entries, due to being exposed to diverse vocabulary items in their daily lives in Japan. In addition, due to their higher proficiency, they were likely to have been sensitive to novel words even at the first occurrence. Therefore, the probability that participants would be able to learn the TWs from a pre-test was high. By the same token, using pseudo-words appeared unethical, as they also would be likely to be learned immediately and remain in their memory.

Second, the vocabulary post-tests were administered to tap into participants' knowledge of newly learned lexical items. Testing instruments for both forms and meaning of the items were carefully designed to use contextual information comparable to what participants had heard in the listening passages. As described in the literature review, how best to assess IVL outcomes, and what types of instrument can most effectively measure newly learned vocabulary knowledge, is still open to question.

Different measures, such as reading times in eye-tracking paradigms, are already used in studies that attempt to assess cognitive processing during learning in the written modality. Most listening research still lacks measures comparable to those used in reading research. Learning outcomes could be measured in more sensitive ways, for example, by analyzing reaction times. Test instruments used to measure intentional learning are inadequate to assess the gain from incidental learning. Further exploration to identify more adequate instruments for incidental learning would be worthwhile. Additionally, despite careful planning of the order of test administration, final MDT results might have included testing effects, due to additional exposures to TWs. Furthermore, it remains unclear whether knowledge obtained from enhanced IVL lasts, and whether it can be retrieved when necessary, questions of interest from the pedagogical perspective. Further exploration of enhanced IVL that includes delayed post-tests is necessary.

Third, the test format of the CQs was designed to draw participants' attention to the content without taxing their WM resources. However, it might have added an extra level of difficulty by presenting each CQ's multiple choices auditorily only. Another option would have been to ask the participants to respond to questions of veracity based on the content of the talks, which might tap into listening comprehension more directly than the multiple choice questions did.

Fourth, results for comprehension questions were different from what was predicted. Specifically, elaborated input did not show significant effects on responses to inference items. As Yano et al. (1994) suggested, different item types require different cognitive processes, and apparent differences by item type were found in conjunction with WM. This also suggests the reason for the low reliability values of comprehension questions in both studies. The modality difference might have resulted in the different outcomes. To obtain a clearer picture of these cognitive processes, further investigations could consider (a) relationships between inference items and input modification type, and (b) relationships between specific question items and different components of WM.

Fifth, due to the resources available and time constraints on the treatment sessions, only two non-verbal complex WM tasks were employed. It would be preferable to include a simple WM task, such as a phonological short-term memory (PSTM) task. Linck et al. (2014) reported verbal WM tasks as better predictors. It would be ideal to include simple and complex verbal and non-verbal tasks in future research.

Finally, it is obvious that more research should be conducted on input modification, especially input elaboration, including modified elaborated input that consists of longer sentences than simplified input. This study used only the auditory mode, but as Long (2019) pointed out, possible effects with bimodality should be explored in the future (see Borro, 2020, and Malone, 2018 for two examples). Bimodality could provide further unobtrusive interventions for incidental learning conditions. Derwing (1996) discussed a future agenda for this line of research, which is to uncover what type of elaboration is most effective and efficient for comprehension and L2 learning. Relationships between input, comprehension, and L2 language acquisition are fundamental elements of SLA, and how input modification contributes to SLA is one of the keys to building theoretical frameworks (Chaudron, 1985b; Long, 1983a, 1985). More research on relationships between comprehension and IVL conditions is needed to extend these fundamental aspects of SLA, to better understand learning as a by-product of IVL, and to overcome the difficulty of operationalizing learners' attention and awareness during learning.

Appendix A. Guidelines for mini-lecture draft

ミニレクチャー原稿のガイドライン

この度は、ご専門に関するミニレクチャーをお願いさせていただき、本当にありがとうございます。原稿を準備される時に参考になるガイドラインを以下にまとめましたので、ご覧ください。

1. このミニレクチャーは、大学進学前の日本人高校生が対象です。今後の専門学部の選択のために様々な学術分野で行われている研究をショーケース的に説明・紹介するセッションで、ご自身の専門分野について、ご説明するという設定でそのお話の原稿をお願いします。他の研究分野の方々も、同席しているとお考えください。
2. ミニレクチャーの制限時間は、4 分です。（NHK のラジオニュースの速度参考）
3. 原稿の字数は、１５００～２０００字くらいとお考えください。
4. パワーポイントスライドなどは使わず、口頭のみでの説明です。
5. 内容は、その専門分野の重要性、課題などを、わかりやすい構成でまとめてください。

   具体例や、データなども大学生が理解できる範囲でお使いください。専門用語や難度の高い語彙が少々入っても構いませんが、専門のことは全く知らない人が聴衆とお考えください。
6. 始めは、簡単な自己紹介からお始めください。

その他、ご質問等ございましたら、私の方へ、直接ご連絡ください。メールでも電話でもどちらでも大丈夫です。（下記参照）

これからホリデーシーズンに向かうお忙しい時期に大変恐縮ですが、この原稿は、今から 2、３週間以内にお願いできませんでしょうか。ホリデーシーズンが終わってからならできるという場合は、是非ご一報ください。

原稿料として心ばかりではありますが、謝礼をご準備させていただきます。

これは、私が現在進めている第二言語習得学と日本語教育学に関する博士論文の一部です。

ご協力とご理解に、心から感謝申し上げます。

突然のことで、大変恐縮ですが、どうぞよろしくお願い申し上げます。

メリーランド大学カレッジパーク校

第二言語習得博士課程

ヒルマン小林恭子

kkh1225@umd.edu, kyokokh@gmail.com

Appendix A. (cont'd). Guidelines for mini-lecture draft: English translation
12/5/2018 Revised

(These guidelines were distributed to potential speakers who are researchers/professors, who have a master's degree or higher, and who could give this type of lecture in real life.)

Thank you for agreeing to provide a mini-lecture about your academic specialty.
Please see below for guidelines to use as a reference when preparing the draft of the lecture.

1. This mini-lecture is for Japanese high school students who will enter university in the near future. Please prepare a draft for a talk, assuming that you will explain your academic field of specialty at a session that showcases introductions to ongoing studies in various academic fields. Please imagine that there will be other researchers from other fields in the same session.
2. The time limit of the mini-lecture is 4 minutes. (The reference speed is radio news on NHK.)
3. The number of characters in the draft should be approximately 1500 to 2000 characters.
4. Please plan to present only orally, without using PowerPoint slides.
5. Please make the lecture content about the significance and main issues of the field easy to follow.
Please use specific examples, data, and so on at a level the student audience can understand. It is fine to include some specialized and difficult vocabulary, but please assume the audience knows nothing about your specialty.
6. Please begin with a brief self-introduction.

Please contact me directly if you have other questions. Either by email or phone is fine (see below).
I am very sorry to request this during this busy pre-holiday season, but could you please write this draft within the next two or three weeks? Please contact me in the event you cannot write it until after the holiday season.

I will prepare a small gift in compensation for the draft as a token of my gratitude.
The drafts will become part of my dissertation project that I am currently working on in the fields of Second Language Acquisition and Japanese Language Education.
Thank you from the bottom of my heart for your cooperation and understanding.
I apologize for the abrupt nature of this request, and I thank you in advance for your work.

Kyoko Kobayashi
HillmanUniversity of Maryland, College Park
kkh1225@umd.edu,
kyokokh@gmail.com

Appendix B. List of target words (TWs) and their frequency

Talk No.		Noun	English translation	Frequency rank: in JFL/JSL materials		Frequency rank: general	
1	1	*shinkutanku* シンクタンク	think-tank	N/A	N/A	N/A	N/A
	2	*hasshoo* 発祥	beginning	10K	9624	11K	10203
	3	*minkan sekutaa* 民間セクター	private sector	N/A	N/A	N/A	N/A
	4	*intaan* インターン	intern	21K+	27875	21K+	29015
	5	*ukezara* 受け皿	receiver	13K	12758	15K	14205
	6	*inhura* インフラ	infrastructure	12K	11173	15K	14425
	7	*enujiioo* NGO (non-government organization)	NGO	N/A	N/A	N/A	N/A
	8	*zesei* 是正	correction	10K	9119	12K	11788
2	1	*honyuurui* 哺乳類	mammal	N/A	N/A	N/A	N/A
	2	*seechuu* 成虫	imago, adult	21K+	28489	21K+	21906
	3	*minkan sekutaa* ボウフラ	wiggler	21K+	34401	21K+	30397
	4	*seesoku* 棲息	inhabitation	21K+	20594	21K+	23322
	5	*jueki* 樹液	tree sap	21K+	25319	21K+	28781
	6	*satoyama* 里山	woodlands near town	N/A	N/A	N/A	N/A
	7	*kyuumitsu* 吸蜜	nectar sucking	N/A	N/A	N/A	N/A
	8	*yasoo* 野草	wild grass	21K+	20548	21K+	24578

Appendix B. (cont'd)

Talk No.	rank	Noun	English translation	Frequency rank in JFL/JSL materials		Frequency rank general	
3	1	*booenkyoo* 望遠鏡	telescope	N/A	N/A	N/A	N/A
	2	*kashikoo* 可視光	visible light	N/A	N/A	N/A	N/A
	3	*ginga(kee)* 銀河(系)	galactic system	15K	14776	17K	16417
	4	*taiko* 太古	primeval time	12K	11529	13K	11554
	5	*eniguma* エニグマ	enigma	N/A	N/A	N/A	N/A
	6	*hooden* 放電	electric discharge	21K+	21695	19K	21721
	7	*baruun* バルーン	air balloon	21K+	32039	21K+	27231
	8	*ryuushi* 粒子	particle	8K	7809	10K	9401
4	1	*kokuban* 黒板	blackboard	12K	11990	13K	12353
	2	*manabiya* 学び舎	school building	N/A	N/A	N/A	N/A
	3	*gekoo* 下校	going home from school	21K+	24566	21K+	23150
	4	*rogo* ロゴ	logo type	12K	11051	11K	10263
	5	*bassui* 抜粋	excerpt	11K	10320	11K	10065
	6	*doriru* ドリル	a notebook for math and kanji practice	12K	11323	12K	11665
	7	*tatewari* 縦割り	hierarchical structure	20K	19640	21K+	19615
	8	*suishoo* 推奨	recommendation	12K	11051	11K	10263

Appendix C. Sample of three types of modification

Genuine

G	Genuine
1	岡澤孝雄と申します。
	(I am) called Takao Okazawa.
2	長い間、血を吸う昆虫の研究をしてきましたが、今日は、血を吸わない蚊の研究についてお話しします。
	For a long time, I have been studying insects that suck blood, but today, I will talk about research on mosquitoes that do not suck blood.
3	血を吸う**昆虫**といっても色々あるわけですが、シラミ、ノミ、蚊などは代表的なもので、**哺乳類**(1)の血を吸って生きています。
	Speaking of insects that suck blood, they vary, but (those) such as lice, fleas, and mosquitos are typical, and they live by sucking the blood of **mammals** (1).
4	これらは血を吸うという行動を通してヒトや動物の病気を媒介しますので私たち人間にとっては大変重要な生物です。
	They are very important creatures for us humans, because they transmit diseases to humans and animals through their behavior of sucking blood.
5	では何故シラミ、ノミ、カは血を吸うのでしょうか。
	Then why do lice, fleas, and mosquitoes suck blood?
6	シラミは幼虫も**成虫**(1)も雌雄に関わらず**哺乳類**(2)の皮膚に住んで、血を吸います。
	As for lice, both larvae and **imago** (1) do suck blood while living on **mammals**' (2) skin regardless of the gender.
7	ノミの幼虫は自由生活をしており血を吸いませんが、**成虫**(2)は雌も雄も吸血します。
	Flea larvae live a free life and do not suck blood, but both male and female **imago** (2) suck blood.
8	シラミ、ノミは餌として血を摂取しています。
	Both lice and fleas take blood as food.
9	カは**成虫**(3)の雌だけが血を吸いますが、これは、血を吸わなければ卵ができないからです。
	As for mosquitoes, only female **imago** (3) suck blood, but this is because they cannot lay eggs without sucking blood.

G	Genuine
10	一方、**ボウフラ**(1)は水中に棲んでおり、血は吸いません。
	On the other hand, **wigglers** (1) live in the water and do not suck blood.
11	**ボウフラ**(2)は水溜りという小生態系の中にいて、最近の研究でわかったのは、他種の幼虫を餌として食べる捕食者の蚊の幼虫がいることです。
	Wigglers (2) are in a small ecological system called a water puddle, and what was found in recent research is that there are larvae of mosquitoes, as predators, that eat other kinds of larvae as food.
12	また**ボウフラ**(3)は、水中では魚などの餌生物でもあります。
	Also, **wigglers** (3) serve as a food source for fish and so on in the water.
13	カは私達の周りに**棲息**(1)が確認できる、身近な昆虫です。
	Mosquitoes are familiar insects that can be studied as their **living** (1) around us.
14	日本にも戦前まではマラリアやフィラリアというカが媒介する病気がありました。
	In Japan as well, there were diseases that were transmitted by mosquitoes, such as malaria and filaria until WWII.
15	1970 年代まではウイルスの病気、日本脳炎も流行していました。
	Until the 1970s, Japanese encephalitis, a viral disease, prevailed.
16	最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告され、国内での二次感染を防ぐための対策を考える必要がでてきました。
	Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well, and strategies to prevent secondary infection domestically are becoming necessary.
17	日本には 130 種ほどの蚊の**棲息**(2)がわかっていますが、ほとんどは吸血性で、**哺乳類**(3)、鳥類、爬虫類、両生類の血を吸います。
	In Japan, it is found the **living** (3) of approximately 130 kinds of mosquitoes, but most of them are hematophagous, and they suck the blood of **mammals** (3), fowls, reptiles, and amphibians.
18	魚を吸血するカもいて、蚊の種類によりどの動物の血を吸うかが決まっていますが、自然に**棲息**(3)が分かっていても、カの吸血行動はまだまだ分からない部分が多いのです。

G	Genuine
	There are also mosquitoes that suck the blood of fish, and determining which animal's blood is sucked depends on the kind of mosquito, but even though their **living** (3) in nature is understood, there are still many unknown parts of mosquitoes' hematophagous behaviors.
19	カは人や家畜の病気を媒介しますから研究者も多く、特に媒介蚊については形態、生理、生態などかなり詳しく研究されてきました。
	Since mosquitoes transmit the diseases of humans and livestock, there are many researchers, who study them, and infection-transmitting mosquitoes especially have had their morphology, physiology, and ecology studied in detail.
20	それではなぜ血を吸わないカの研究をするのか？
	Then, why do I conduct research on mosquitoes that do not suck blood?
21	理由の一つは、吸血する雌と同じぐらいの数がいるはずの雄についてどのような生活をしているのかあまり研究がなく、それが分かれば雄をターゲットとした蚊の防除法が開発されるかもしれないからです。
	One reason is that there is not much research on how males, which are supposed to be as numerous as hematophagous females, live, and if this is understood, pest control methods targeting males may be developed.
22	カの雄が花の蜜を吸うことや植物の傷から**樹液**(1)を吸う断片的な報告がありますが、どの蚊がどの花に集まるのか野外で調べることは難しいのです。
	There are fragmentary reports about male mosquitoes' eating flower nectar and sucking **tree sap** (1) from weak spots in plants, but it is difficult to investigate which mosquitoes gather at which flowers in the field.
23	それで、昨年 9 月に金沢の里山(1)で私 が行った研究についてお話ししたいと思います。
	So, I would like to talk about a study that I conducted in the **woodlands** (1) of Kanazawa last September.
24	その時、3 種類のカが**吸蜜**(1)をしているのを見つけました。
	I found three kinds of mosquitoes doing **feeding on nectar** (1).
25	それは、9 種類の**野草**(1)の花のところで、その中でミズヒキという花には、キンパラナガハシカというカの雄が沢山来ていました。
	That was on the flowers of nine kinds of **wild grass** (1) where a lot of male mosquitoes, called *kinpara nagahashika*, came to the flower called *mizuhiki*.

G	Genuine
26	このことは20年前に私が富山で観察して分かっていたものの、金沢大学に移って20年、毎年**樹液**(2)があるところや、**野草**(2)のミズヒキの開花期に里山(2)に行って観察していて、金沢ではキンパラナガハシカを見つけることができませんでした。
	Although this became known to me when I conducted observations in Toyama 20 years ago, it has been 20 years since I moved to Kanazawa University, and I went to the **woodlands** (2) every year to make observations where **tree sap** (2) is present, and at the time the **wild grass** (2) *mizuhiki* flowers begin to bloom, but I could not find *kinpara nagahashika* in Kanazawa.
27	昨年20年ぶりにキンパラナガハシカをミズヒキの花で見つけ、詳しく観察しました。
	Last year, after 20 years, I found *kinpara nagahashika* mosquitoes on the *mizuhiki* flowers, and observed (them) in detail.
28	その結果、ミズヒキの花は朝9時頃から12時にかけて満開になり、それに合わせてキンパラナガハシカの雄が午前中だけ**吸蜜**(2)に来る事が分かりました。
	The results were that because *mizuhiki* flowers come into full bloom in the morning from 9:00 through 12:00, it is in response to this that male *kinpara nagahashika* mosquitoes come to do their **feeding on nectar** (2) only in the morning.
29	富山ではいつも午前中に観察し、金沢の里山(3)では昨年以前は午後に観察していたという単純なことが結果を左右していました。
	The simple fact that I had always observed in the morning in Toyama while I had observed in the afternoon in the **woodlands** (3) of Kanazawa before last year, affected the results.
30	ミズヒキとは対照的にヒヨドリバナには午後2時ごろから5時ごろに蚊が**吸蜜**(3)に来ます。
	In contrast to *mizuhiki*, mosquitoes come to *hiyodori bana* flowers for their **feeding on nectar** (3) between 2:00 and 5:00 in the afternoon.
31	カは自然の中で季節ごとに変わる植物の開花に合わせ利用する花を変え、また一日の中でも時間帯によって利用する花を変えることが明らかになりました。

G	Genuine
	It became clear that mosquitoes in nature change the flowers that they use in response to plants' full bloom periods that change every season, and that they change the flowers that they use depending on the time of day.
32	人にとってカは病気を媒介する害虫として有名ですが、実は、カはこれらの**野草**(3)の花粉媒介者でもあり、血を吸わない雄が植物の花粉を媒介すると聞いて驚く人もいるかもしれません。
	Although mosquitoes are notorious as pests that transmit diseases, in fact, the mosquitoes are pollinators of this **wild grass** (3), and there may be people who are surprised to hear that males that do not suck blood but transfer plant pollen.
33	また、将来**樹液**(3)を吸う蚊も観察できるかもしれません。
	Also, we may be able to observe mosquitoes that suck **tree sap** (3) in the future.
34	このように、世界で 3500 種いると言われるカ類の生態系の中での役割の理解はまだまだ進んでおらず、これから面白い発見が色々出て来ると思います。
	As mentioned here, understanding of the roles in the ecological system of the mosquito species of which there are said to be 3500 types in the world is not yet advanced, so I think that a variety of interesting findings will continue to emerge from now on.

Simplified (G: Genuine; S: Sub-sentences [in genuine]; T: Total sentences)

G	S	T	Simplified
1	1	1	岡澤孝雄と言います。
			(I am) said Takao Okazawa.
2	1	2	長い間、血を吸う昆虫の研究をしてきました。
			For a long time, I have been studying insects that suck blood.
	2	3	しかし、今日は、血を吸わない蚊の研究についてお話しします。
			However, today, I will talk about research on mosquitoes that do not suck blood.
3	1	4	血を吸う昆虫といっても色々あるわけです。
			Speaking of insects that suck blood, they vary.
	2	5	シラミ、ノミ、蚊などは代表的なもので、**哺乳類**(1)の血を吸って生きています。
			(Those) such as lice, fleas, and mosquitoes are typical, and they live by sucking the blood of **mammals** (1).
4	1	6	これらは血を吸うという行動を通してヒトや動物の病気を広く移します。
			These (insects) widely transmit diseases to humans and animals through their behavior of sucking blood.
	2	7	それで、私たち人間にとっては大変重要な生物です。
			Therefore, they are very important creatures for us humans.
5	1	8	では何故シラミ、ノミ、カは血を吸うのでしょうか。
			Then why do lice, fleas, and mosquitoes suck blood?
6	1	9	シラミは子供の時も**成虫**(1)も雄と雌に関わらず、**哺乳類**(2)の体に住んで、血を吸います。
			As for lice, they suck blood both during childhood and when **imago** (1), while living on **mammals**' (2) bodies, whether female or male.
7	1	10	ノミの子供は自由生活をしており血を吸いません。
			Flea larvae live a free life and do not suck blood.
	2	11	**成虫**(2)は雌も雄も血を吸います。
			Both male and female **imago** (2) suck blood.
8	1	12	シラミ、ノミは餌として血を吸っています。
			Both lice and fleas suck blood as food.
9	1	13	カは**成虫**(3)の雌だけが血を吸います。
			As for mosquitoes, only female **imago** (3) suck blood.

G	S	T	Simplified
	2	14	これは、血を吸わなければ卵ができないからです。
			This is because they cannot lay eggs without sucking blood.
10	1	15	一方、ボウフラ(1)は水の中に棲んでいます。
			On the other hand, **wigglers** (1) live in the water.
	2	16	血は吸いません。
			They do not suck blood.
11	1	17	ボウフラ(2)は水溜りという小さな生き物の世界の中にいます。
			Wigglers (2) are in a world of small creatures called a water puddle.
	2	18	他の種類の子供を餌として食べる蚊の子供もいます。
			There are children of mosquitoes that eat children of other kinds of insects as food.
	3	19	最近の研究でわかりました。
			This was found in recent research.
12	1	20	またボウフラ(3)は、水中では魚などのエサでもあります。
			Also, **wigglers** (3) serve as food for fish and others in the water.
13	1	21	カは私達の周りに棲息(1)が確認できます。
			Mosquitoes can be checked as their **living** (1) around us.
	2	22	身近な昆虫です。
			They are familiar insects.
14	1	23	日本にも戦前まではマラリアやフィラリアというカが移す病気がありました。
			In Japan as well, there were diseases that were spread by mosquitoes, such as malaria and filaria until WWII.
15	1	24	1970 年代まではウイルスの病気、日本脳炎も流行していました。
			Until the 1970s, Japanese encephalitis, a viral disease, prevailed.
16	1	25	最近、世界の暑い国の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告されました。
			Recently, patients with dengue fever and zika fever, which were thought to be diseases from the hot countries of the world, have been reported in Japan as well.
	2	26	それで、国内での人から人へ移るのを防ぐための対策を考える必要がでてきました。

G	S	T	Simplified
			Therefore, strategies to prevent getting them passed from one person to another domestically are becoming necessary.
17	1	27	日本には 130 種ほどの蚊の**棲息**(2)がわかっています。
			In Japan, it is found the **living** (2) of approximately 130 kinds of mosquitoes
	2	28	しかし、ほとんどは血を吸う種類で、**哺乳類**(3)、鳥、ヘビやとかげ、カエルなどの血を吸います。
			However, most of them are kinds that suck blood, and they suck the blood of **mammals** (3), birds, snakes, lizards, frogs and so on.
18	1	29	魚の血を吸うカもいます。
			There are also mosquitoes that suck the blood of fish.
	2	30	蚊の種類によりどの動物の血を吸うかが決まっています。
			Determining which animal's blood is sucked depends on the kind of mosquito.
	3	31	しかし、自然に**棲息**(3)が分かっていても、カの血を吸う行動はまだまだ分からない部分が多いのです。
			However, even though their **living** (3) in nature is understood, there are still many unknown parts of mosquitoes' behaviors of sucking blood.
19	1	32	カは人や人といっしょにいる動物の病気を移します。
			Mosquitoes pass the diseases of humans and animals that live with humans.
	2	33	ですから研究者も多いです。
			Therefore, there are many researchers on mosquitos
	3	34	特に病気を移す蚊については形、働き、生きている環境などかなり詳しく研究されてきました。
			Disease-passing mosquitoes especially have had their shapes, roles, and living environment studied in detail.
20	1	35	それではなぜ血を吸わないカの研究をするのか？
			Then, why do I conduct research on mosquitoes that do not suck blood?
21	1	36	理由の一つは、血を吸う雌と同じぐらいの数がいるはずの雄についてどのような生活をしているのかあまり研究がありませんでした。
			One reason is that there was not much research on how males, which are supposed to be as numerous as the blood-sucking females, live.

G	S	T	Simplified
	2	37	それが分かれば雄をターゲットとした蚊の減らし方が開発されるかもしれません。
			If this is understood, reduction methods targeting males may be developed.
22	1	38	カの雄が花の蜜を吸うことや植物の傷から**樹液**(1) を吸う少ない報告があります。
			There are few reports about male mosquitoes' eating of flower nectar and sucking of **tree sap** (1) from weak spots in plants.
	2	39	しかし、どの蚊がどの花に集まるのか野外で調べることは難しいのです。
			However, it is difficult to investigate which mosquitoes gather at which flowers in the field.
23	1	40	それで、昨年 9 月に金沢の里山(1)で私の行った研究についてお話ししたいと思います。
			So, I would like to talk about a study I conducted in the **woodlands** (1) of Kanazawa last September
24			その時、3 種類のカが**吸蜜**(1)をしているのを見つけました。
			I found three kinds of mosquitoes **feeding on nectar** (1).
25	1	41	それは、9 種類の**野草**(1)の花のところでした。
			That was on the flowers of nine kinds of **wild grass** (1).
	2	42	その中でミズヒキという花には、キンパラナガハシカというカの雄が沢山来ていました。
			It was here where a lot of male mosquitoes, called *kinpara nagahashika*, came to the flower called *mizuhiki*.
26	1	43	このことは 20 年前に私が富山で観察して分かっていました。
			This became known to me when I conducted observations in Toyama 20 years ago.
	2	44	金沢大学に移って 20 年、毎年**樹液**(2)があるところや、**野草**(2)のミズヒキの花が咲くころに里山(2)に行って観察していました。
			It has been 20 years since I moved to Kanazawa University, and I went to the **woodlands (2)** to make observations every year at a place where there is **tree sap** (2) and at the time the **wild grass** (2) *mizuhiki* flowers begin to bloom.

G	S	T	Simplified
	3	45	しかし、金沢ではキンパラナガハシカを見つけることができませんでした。
			However, I could not find *kinpara nagahashika* in Kanazawa.
27	1	46	昨年20年ぶりにキンパラナガハシカをミズヒキの花で見つけました。
			Last year, after 20 years, I found *kinpara nagahashika* mosquitoes on the *mizuhiki* flowers.
	2	47	そして、詳しく観察しました。
			And (I) observed (them) in detail.
28	1	48	その結果、ミズヒキの花は朝9時頃から12時にかけて満開になることがわかりました。
			The results were that *mizuhiki* flowers come into full bloom in the morning from 9:00 through 12:00.
	2	49	そして、それに合わせてキンパラナガハシカの雄が午前中だけ**吸蜜**(2)に来る事が分かりました。
			And it was found that in response to this, male *kinpara nagahashika* mosquitoes came for **feeding on nectar** (2) only in the morning.
29	1	50	富山ではいつも午前中に観察していました。
			I had always observed in the morning in Toyama.
	2	51	金沢の里山(3)では昨年以前は午後に観察していました。
			In the **woodlands (3)** of Kanazawa, I had observed in the afternoon before last year.
	3	52	そんな単純なことが結果を左右していました。
			Such a simple thing affected the results.
30	1	53	ミズヒキとは反対にヒヨドリバナには午後2時ごろから5時ごろに蚊が**吸蜜**(3)に来ます。
			In contrast to *mizuhiki*, mosquitoes come to *hiyodori bana* flowers for **feeding on nectar** (3) between 2:00 and 5:00 in the afternoon.
31	1	54	自然の中で季節ごとに花の咲く時期が変わります。
			In nature, plants' full bloom periods change every season.
	2	55	カはそれに合わせて、利用する花を変えることがわかりました。
			It was found that mosquitoes adjust to this accordingly, and that they change the flowers they use.

G	S	T	Simplified
	3	56	また一日の中でも時間帯によって利用する花を変えることがわかりました。
			Also, it was found that based on the time of day, they (mosquitoes) change the flowers they use.
32	1	57	人にとってカは病気を移す悪い虫として有名です。
			Mosquitoes are notorious as bad insects that pass diseases.
	2	58	でも、実はカはこれらの**野草**(3)の花粉を運ぶものでもあります。
			But, in fact, the mosquitoes are also carriers of the plant pollen of this **wild grass** (3).
	3	59	血を吸わない雄が植物の花粉を運ぶと聞いて驚く人もいるかもしれません。
			There may be people who are surprised to hear that males, which do not suck blood, transfer the plant pollen.
33	1	60	また、将来**樹液**(3)を吸う蚊も観察できるかもしれません。
			Also, we may be able to observe mosquitoes that suck **tree sap** (3) in the future.
34	1	61	カ類は、世界で 3500 種いると言われています。
			It is said that there are 3500 types of mosquito species in the world.
	2	62	しかし、このように、カの生きている環境の中での役割の理解はまだまだ進んでいません。
			However, as mentioned here, understanding of their roles in the environment in which the mosquitoes are living is not yet advanced.
	3	63	これから面白い発見が色々出て来ると思います。
			I think that a variety of interesting findings will emerge from now on.

Elaborated (G: Genuine; S: Sub-sentences [in genuine]; T: Total sentences)

G	S	T	Elaborated
1	1	1	岡澤孝雄と申します。
			(I am) called Takao Okazawa.
2	1	2	長い間、血を吸う昆虫、つまり、虫、の研究をしてきましたが、今日は、血を吸わない蚊の研究についてお話しします。
			For a long time, I have been studying insects, that is insects that suck blood, however, today, I will talk about research on mosquitoes that do not suck blood.
3	1	3	血を吸う昆虫といっても色々あるわけですが、シラミ、ノミ、蚊などは代表的なもので、**哺乳類**(1)、つまり、人や動物などのグループ、の血を吸って生きています。
			Speaking of insects that suck blood, they vary, but (those) such as lice, fleas, and mosquitos are typical, and they live by sucking the blood of **mammals** (1), or the group that includes humans and animals.
4	1	4	これら、シラミ、ノミ、カなどは血を吸うという行動を通して、ヒトや動物の病気を媒介しますので、私たち人間にとっては大変重要な生物です。
			They are very important creatures for us humans because these (insects) such as lice, fleas, and mosquitoes transmit diseases to humans and animals through their behavior of sucking blood.
	2	5	言い換えると、シラミ、ノミ、カなどが血を吸って、ヒトや動物を病気にすることがわかっているので、その研究が大切です。
			In other words, because we know that (insects) such as lice, fleas, and mosquitoes make people sick when they suck blood, research on them is very important.
5	1	6	では何故シラミ、ノミ、カは血を吸うのでしょうか。
			Then why do lice, fleas, and mosquitoes suck blood?
6	1	7	シラミは幼虫、つまり虫の子供も**成虫**(1)、つまり、十分大きくなって、もう形が変わらない状態の虫、も雌雄に関わらず、雌であっても雄であっても**哺乳類**(2)の皮膚に住んで、血を吸います。

G	S	T	Elaborated
			As for lice, both larvae, or young insects, and **imago** (1), or insects that are big enough to have reached a stage where their shape does not change, suck blood while living on **mammals**' (2) skin regardless of gender, whether they are female or male (lice).
7	1	8	ノミの幼虫は自由生活をしており血を吸いません。
			Flea larvae live a free life and do not suck blood.
	2	9	**成虫**(2)は雌も雄も吸血します。
			Both male and female **imago** (2) do suck blood.
8	1	10	シラミ、ノミは餌として血を摂取しています。
			Both lice and fleas take blood as food.
	2	11	血が栄養となっているということです。
			In other words, blood is their source of nutrition.
9	1	12	カは**成虫**(3)の雌だけが血を吸います。
			As for mosquitoes, only female **imago** (3) suck blood.
	2	13	これは、血を吸わなければ卵ができないからです。
			This is because they cannot lay eggs without sucking blood.
10	1	14	一方、**ボウフラ**(1)、つまり、蚊が大きくなる前の幼虫、は水中に棲んでおり、血は吸いません。
			On the other hand, **wigglers** (1), or mosquito larvae before they become big, live in the water and do not suck blood.
11	1	15	**ボウフラ**(2)は水溜りという小生態系の中にいて、最近の研究でわかったのは、他種の幼虫を餌として食べる捕食者の蚊の幼虫がいることです。
			Wigglers (2) are in a small ecological system called a water puddle, and what was found in recent research is that there are mosquito larvae that as predators eat the larvae of other kinds as food.
12	1	16	また**ボウフラ**(3)は、水中では魚などの餌生物でもあります。
			Also, **wigglers** (3) serve as a food source for fish and others in the water.
	2	17	言い換えると、それらは、ほかの幼虫を食べたり、魚などに食べられたりして、水たまりのような小さい生態系の中で生きています。

G	S	T	Elaborated
			In other words, those eat other larvae, and are eaten by (others) such as fish, and they live inside a small ecological system like a water puddle.
13	1	18	カは私達の周りに**棲息**(1)、つまり、住んで生きていることが確認できる、身近な昆虫です。
			Mosquitoes are familiar insects that can be checked as their **living** (3) around us, or their status of living nearby and being alive.
14	1	19	日本にも戦前まではマラリアやフィラリアというカが媒介する病気がありました。
			In Japan as well, there were diseases that were transmitted by mosquitoes such as malaria and filaria until WWII.
15	1	20	1970年代まではウイルスの病気、日本脳炎も流行していました。
			Until the 1970s, Japanese encephalitis, a viral disease, prevailed.
16	1	21	最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告され、国内での二次感染を防ぐための対策を考える必要がでてきました。
			Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well, and strategies to prevent domestic secondary infection are becoming necessary.
	2	22	言い換えると、日本は熱帯ではありませんが、デング熱、ジカ熱などの患者が日本でも報告されているので、その患者から他の人に移らないようにする必要が出て来ました。
			In other words, Japan is not in the tropics, but patients with diseases, such as dengue fever and zika fever have been reported in Japan; therefore, in Japan it is becoming necessary to make sure these diseases do not get passed from these patients to others.
17	1	23	日本には130種ほどの蚊の**棲息**(3)がわかっていますが、ほとんどは吸血性で、哺乳類(3)、トリなどの鳥類、ヘビやトカゲなどの爬虫類、カエルなどの両生類の血を吸います。

G	S	T	Elaborated
			In Japan, it is found the **living** (3) of approximately 130 kinds of mosquitoes, but most of them are hematophagous, and they suck the blood of **mammals** (3), fowls, such as chickens, reptiles, such as snakes and lizards, and amphibians, such as frogs.
18	1	24	魚を吸血するカもいて、蚊の種類によりどの動物の血を吸うかが決まっていますが、自然に**棲息**(3)が分かっていても、カの吸血行動、つまり、どのように血を吸っているか、はまだまだ分からない部分が多いのです。
			There are also mosquitoes that suck the blood of fish, and determining which animal's blood is sucked depends on the kind of mosquito, but even though their **living** (3) in nature is understood, there are still many unknown parts of mosquitoes' hematophagous behaviors, or how they suck blood.
19	1	25	カは人や家畜の病気を媒介しますから研究者も多く、特に媒介蚊については形態、生理、生態などかなり詳しく研究されてきました。
			Since mosquitoes transmit the diseases of humans and livestock, there are many researchers who study them, and infection-transmitting mosquitoes especially have had their morphology, physiology, and ecology studied in detail.
	2	26	言い換えると、多くの研究者が、ヒトや牛、馬などの家畜の病気を移すカの形や、出るところなどについてかなり研究してきたということです。
			In other words, many researchers have studied in detail the shapes of mosquitoes that transmit diseases of humans and livestock like cows and horses, and places where (the mosquitoes) appear.
20	1	27	それではなぜ血を吸わないカの研究をするのか。
			Then, why do I conduct research on mosquitoes that do not suck blood?
21	1	28	理由の一つは、吸血する雌と同じぐらいの数がいるはずの雄についてどのような生活をしているのかあまり研究がありませんでした。
			One reason is that there was not much research on how males, which are supposed to be as numerous as hematophagous females, live.

G	S	T	Elaborated
	2	29	それが分かれば雄をターゲットとした蚊の防除法、減らし方が開発されるかもしれません。
			If this is understood, pest control methods targeting males, or reduction methods, may be developed.
22	1	30	カの雄が花の蜜を吸うことや植物の傷から**樹液**(1)、つまり、木の幹から出てくる水のような液体を吸う断片的な、少しの部分的な報告がありますが、どの蚊がどの花に集まるのか野外で調べることは難しいのです。
			There are limited, partial, fragmentary reports about male mosquitoes' eating flower nectar and sucking **tree sap** (1), or a liquid like water that comes out of a tree trunk, from weak spots in plants, but it is difficult to investigate which mosquitoes gather on what flowers outside.
23	1	31	それで、昨年 9 月に金沢の里山(1)で私の行った研究についてお話したいと思います。
			So, in the research I conducted, last September in the **woodlands**(1) of Kanazawa.
24			その時、3 種類のカが**吸蜜**(1)、つまり、花の蜜を吸うこと、をしているのを見つけました。
			I found three kinds of mosquitoes **feeding on nectar** (1), or sucking nectar from flowers.
25	1	32	それは、9 種類の**野草**(1)、つまり、山や野原など人が管理していないところに生えている草、の花のところで、その中でミズヒキという花には、キンパラナガハシカというカの雄が沢山来ていました。
			That was on the flowers of nine kinds of **wild grass** (1), or grass growing in the mountains and wild fields not maintained by humans, where a lot of male mosquitoes, called *kinpara nagahashika*, came to the flower called *mizuhiki*.
26	1	33	このことは 20 年前に私が富山で観察して分かっていました。
			This became known to me when I conducted observations in Toyama 20 years ago.

G	S	T	Elaborated
	2	34	私は富山から金沢大学に移って 20 年たち、金沢でも毎年**樹液**(2)があるところや、**野草**(2)のミズヒキの開花期、花が咲くころに**里山**(2)に行って観察していました。
			It has been 20 years since I moved from Toyama to Kanazawa University, and I went to the **woodlands** (2) in Kanazawa to make observations every year at a place where there is **tree sap** (2), and at the time the **wild grass** (2) *mizuhiki* flowers begin to bloom.
	3	35	しかし、金沢では富山のようにキンパラナガハシカを見つけることができませんでした。
			However, I could not find *kinpara nagahashika* in Kanazawa as I did in Toyama.
27	1	36	昨年 20 年ぶりにキンパラナガハシカをミズヒキの花で見つけ、詳しく観察しました。
			Last year, after 20 years, I found *kinpara nagahashika* mosquitoes on the *mizuhiki* flowers, and observed (them) in detail.
28	1	37	その結果、ミズヒキの花は朝 9 時頃から 12 時にかけて満開になり、それに合わせてキンパラナガハシカの雄が午前中だけ**吸蜜**(2)に来る事が分かりました。
			The results were that because *mizuhiki* flowers come into full bloom in the morning from 9:00 through 12:00, it was found that in response to this, male *kinpara nagahashika* mosquitoes came for **feeding on nectar** (2) only in the morning.
29	1	38	富山ではいつも午前中に観察し、金沢の**里山**(3)では昨年以前は午後に観察していたという単純なことが結果を左右していました。
			The simple fact that I had always observed in the morning in Toyama while I had observed in the afternoon in the **woodlands** (3) of Kanazawa before last year affected the results.
	2	39	言い換えると、金沢では午前に観察していなかったから、見つけられなかったということです。
			In other words, the reason I did not find them in Kanazawa was because I was not observing in the morning in Kanazawa.

G	S	T	Elaborated
30	1	40	ミズヒキとは対照的にヒヨドリバナには午後２時ごろから５時ごろに蚊が**吸蜜**(3)に来ます。
			In contrast to *mizuhiki*, mosquitoes come to *hiyodori bana* flowers to do their **nectar** (3) between 2:00 and 5:00 in the afternoon.
31	1	41	このことから、カは自然の中で季節ごとに変わる植物の開花に合わせ利用する花を変え、また一日の中でも時間帯によって利用する花を変えることが明らかになりました。
			From this, it became clear that mosquitoes in nature change the flowers that they use in response to plants' full bloom periods that change every season, and that they change the flowers that they use depending on the time of day.
	2	42	言い換えると、カは季節や時間で蜜を吸う花を変えているということです。
			In other words, mosquitoes change the flowers they suck nectar from based on season and time.
	3	43	カはこれらの**野草**(3)の花粉媒介者、つまり、花粉を運ぶものでもあります。
			Mosquitoes are pollinators of this **wild grass** (3), or carriers of the plant pollen.
32	1	44	人にとってカは病気を媒介する害虫として有名ですが、血を吸わない雄が植物の花粉を媒介すると聞いて驚く人もいるかもしれません。
			Mosquitoes are notorious as pests that transmit diseases, but there may be people who are surprised to hear that males, which do not suck blood, transfer the plant pollen.
33	1	45	また、将来樹液(3)を吸う蚊も観察できるかもしれません。
			Also, we may be able to observe mosquitoes that suck **tree sap** (3) in the future.

G	S	T	Elaborated
34	1	46	このように、カの研究の中では、世界で3500種いると言われるカ類の生態系の中での役割の理解はまだまだ進んでおらず、これから面白い発見が色々出て来ると思います。
			As mentioned here, in mosquito research, understanding of the roles in the ecological system of the mosquito species, of which there is said to be 3500 types in the world, is not yet advanced, so I think that a variety of interesting findings will continue to emerge from now on.

Modified elaborated (G: Genuine; S: Sub-sentences [in genuine]; T: Total sentences)

G	S	T	Modified elaborated
1	1	1	岡澤孝雄と申します。
			(I am) called Takao Okazawa.
2	1	2	長い間、血を吸う**昆虫**、つまり、虫、の研究をしてきました。
			For a long time, I have been studying insects, or bugs, that suck blood.
	2	3	しかし、今日は、血を吸わない蚊の研究についてお話しします。
			However, today, I will talk about research on mosquitoes that do not suck blood.
3	1	4	血を吸う昆虫といっても色々あるわけです。
			Speaking of insects that suck blood, they vary.
	2	5	シラミ、ノミ、蚊などは代表的なもので、**哺乳類**(1)、つまり、人や動物などのグループの血を吸って生きています。
			Those such as lice, fleas, and mosquitos are typical, and they live by sucking the blood of **mammals** (1), or the group that includes humans and animals.
4	1	6	これら、シラミ、ノミ、カなどは血を吸うという行動を通して、ヒトや動物の病気を媒介します。
			These (insects) such as lice, fleas, and mosquitoes transmit diseases to humans and animals through their behavior of sucking blood.
	2	7	それで、私たち人間にとっては大変重要な生物です。
			Therefore, they are very important creatures for us humans.
	3	8	言い換えると、シラミ、ノミ、カなどが血を吸って、ヒトや動物を病気にすることがわかっています。
			In other words, we know that (insects) such as lice, fleas, and mosquitoes make people sick when they suck blood.
	4	9	それで、その研究が大切です。
			Therefore, research on them is very important.
5	1	10	では何故シラミ、ノミ、カは血を吸うのでしょうか。
			Then why do lice, fleas, and mosquitoes suck blood?

G	S	T	Modified elaborated
6	1	11	シラミは幼虫、つまり虫の子供も**成虫**(1)、つまり、十分大きくなって、もう形が変わらない状態の虫、も雌雄に関わらず、雌であっても雄であっても**哺乳類**(2)の皮膚に住んで、血を吸います。
			As for lice, both larvae, or young insects, and **imago** (1), or insects that are big enough to have reached a stage where their shape does not change, suck blood while living on **mammals**’ (2) skin regardless of gender, whether they are female or male (lice).
7	1	12	ノミの幼虫は自由生活をしており血を吸いません。
			Flea larvae live a free life and do not suck blood.
	2	13	**成虫**(2)は雌も雄も吸血します。
			Both male and female **imago** (2) do suck blood.
8	1	14	シラミ、ノミは餌として血を摂取しています。
			Both lice and fleas take blood as food.
	2	15	血が栄養となっているということです。
			In other words, blood is their source of nutrition.
9	1	16	カは**成虫**(3)の雌だけが血を吸います。
			As for mosquitoes, only female **imago** (3) suck blood.
	2	17	これは、血を吸わなければ卵ができないからです。
			This is because they cannot lay eggs without sucking blood.
10	1	18	一方、**ボウフラ**(1)、つまり、蚊が大きくなる前の幼虫、は水中に棲んでいます。
			On the other hand, **wigglers** (1), or mosquito larvae before they become big, live in the water.
	2	19	血は吸いません。
			They do not suck blood.
11	1	20	**ボウフラ**(2)は水溜りという小生態系の中にいます。
			Wigglers (2) are in a small ecological system called a water puddle.
	2	21	他種の幼虫を餌として食べる捕食者の蚊の幼虫もいます。
			There are mosquito larvae that, as predators, eat the larvae of other kinds as food.
	3	22	最近の研究でわかりました。
			This was found through recent research.
12	1	23	また**ボウフラ**(3)は、水中では魚などの餌生物でもあります。
G	S	T	Modified elaborated

			Also, **wigglers** (3) serve as a food source for fish and others in the water.
	2	24	言い換えると、それらは、ほかの幼虫を食べたり、魚などに食べられたりしています。
			In other words, they eat other larvae, and are eaten by (others) such as fish.
	3	25	そして、水たまりのような小さい生態系の中で生きています。
			And they live inside a small ecological system like a water puddle.
13	1	26	カは私達の周りに**棲息**(1)、つまり、住んで生きていること、が確認できます。
			Mosquitoes can be checked as their **living** (1) around us, or their status of living nearby and being alive.
	2	27	身近な昆虫です。
			They are familiar insects.
14	1	28	日本にも戦前まではマラリアやフィラリアというカが媒介する病気がありました。
			In Japan as well, there were diseases that were transmitted by mosquitoes, such as malaria and filaria until WWII.
15	1	29	1970 年代まではウイルスの病気、日本脳炎も流行していました。
			Until the 1970s, Japanese encephalitis, a viral disease, prevailed.
16	1	30	最近、熱帯の病気と考えられていたデング熱、ジカ熱などの患者が日本でも報告されました。
			Recently, patients with dengue fever and zika fever, which were thought to be diseases of tropical areas, have been reported in Japan as well
	2	31	それで、国内での二次感染を防ぐための対策を考える必要がでてきました。
			Therefore, strategies that prevent domestic secondary infection are becoming necessary.
	3	32	言い換えると、日本は熱帯ではありません。
			In other words, Japan is not in the tropics.
	4	33	しかし、デング熱、ジカ熱などの患者が日本でも報告されています。
G	S	T	Modified elaborated
			However, patients with diseases, such as dengue fever and zika

			fever have been reported in Japan as well.
	5	34	それで、その患者から他の人に移らないようにする必要が出て来ました。
			Therefore, it is becoming necessary to make sure these diseases do not get passed from these patients to others.
17	1	35	日本には 130 種ほどの蚊の**棲息**(2)がわかっています。
			In Japan, it is found the **living** (2) of approximately 130 kinds of mosquitoes.
	2	36	しかし、ほとんどは吸血性で、**哺乳類**(3)、トリなどの鳥類、ヘビやトカゲなどの爬虫類、カエルなどの両生類の血を吸います。
			However, most of them are hematophagous, and they suck the blood of **mammals** (3), fowls, such as chickens, reptiles, such as snakes and lizards, and amphibians, such as frogs.
18	1	37	魚を吸血するカもいます。
			There are also mosquitoes that suck the blood of fish.
	2	38	蚊の種類によりどの動物の血を吸うかが決まっています。
			Determining which animal's blood is sucked depends on the kind of mosquito.
	3	39	しかし、自然に**棲息**(3)が分かっていても、カの吸血行動、つまり、どのように血を吸っているか、はまだまだ分からない部分が多いのです。
			However, even though their **living** (3) in nature is understood, there are still many unknown parts of mosquitoes' hematophagous behaviors or how they suck blood.
19	1	40	カは人や家畜の病気を媒介します。
			Mosquitoes transmit the diseases of humans and livestock.
	2	41	ですから研究者も多いです。
			Therefore, there are many researchers (who study them).
	3	42	特に媒介蚊については形態、生理、生態などかなり詳しく研究されてきました。
			Infection-transmitting mosquitoes especially have had their morphology, physiology, and ecology studied in detail.
G	S	T	Modified elaborated
	4	43	言い換えると、多くの研究者が、ヒトや牛、馬などの家畜の病気を移すカの形や、出るところなどについてかなり研究し

			てきたということです。
			In other words, many researchers have studied in detail the shapes of mosquitoes that transmit diseases of humans and livestock like cows and horses, and places where (the mosquitoes) appear.
20	1	44	それではなぜ血を吸わないカの研究をするのか。
			Then, why do I conduct research on mosquitoes that do not suck blood?
21	1	45	理由の一つは、吸血する雌と同じぐらいの数がいるはずの雄についてどのような生活をしているのかあまり研究がありませんでした。
			One reason is that there was not much research on how males, which are supposed to be as numerous as hematophagous females, live.
	2	46	それが分かれば雄をターゲットとした蚊の防除法、減らし方が開発されるかもしれません。
			If this is understood, pest control methods targeting males, or reduction methods, may be developed.
22	1	47	カの雄が花の蜜を吸うことや植物の傷から**樹液**(1)、つまり、木の幹から出てくる水のような液体を吸う断片的な、少しの部分的な報告があります。
			There are limited, partial, fragmentary reports about male mosquitoes' eating flower nectar and sucking **tree sap** (1), or a liquid like water that comes out of a tree trunk, from weak spots in plants.
	2	48	しかし、どの蚊がどの花に集まるのか野外で調べることは難しいのです。
			However, it is difficult to investigate which mosquitoes gather at which flowers in the field.
23	1	49	それで、昨年 9 月に金沢の**里山**(1)で私が行った研究についてお話ししたいと思います。
			So, I would like to talk about a study I conducted in the **woodlands** (1)of Kanazawa last September
24			その時、3 種類のカが**吸蜜**(1)、つまり、花の蜜を吸うこと、をしているのを見つけました。
G	S	T	Modified elaborated
			I found three kinds of mosquitoes that do **feeding on nectar** (1), or sucking nectar from flowers.

25	1	50	それは、9種類の**野草**(1)、つまり、山や野原など人が管理していないところに生えている草、の花のところでした。
			That was on the flowers of nine kinds of **wild grass** (1), or grass growing in the mountains and wild fields not maintained by humans.
	2	51	その中でミズヒキという花には、キンパラナガハシカというカの雄が沢山来ていました。
			It was here where a lot of male mosquitoes called *kinpara nagahashika*, came to the flower called *mizuhiki*.
26	1	52	このことは20年前に私が富山で観察して分かっていました。
			This became known to me when I conducted observations in Toyama 20 years ago.
	2	53	私は富山から金沢大学に移って20年たちます。
			It has been 20 years since I moved from Toyama to Kanazawa University.
	3	54	金沢でも毎年**樹液**(2)があるところや、**野草**(2)のミズヒキの開花期、花が咲くころに**里山**(2)に行って観察していました。
			In Kanazawa, I also went to the **woodlands** (2) to make observations every year, where there is **tree sap** (2) and at the time the **wild grass** (2) *mizuhiki* flowers begin to bloom.
	4	55	しかし金沢では富山のようにキンパラナガハシカを見つけることができませんでした。
			However, I could not find *kinpara nagahashika* in Kanazawa as I did in Toyama.
27	1	56	昨年20年ぶりにキンパラナガハシカをミズヒキの花で見つけました。
			Last year, after 20 years, I found *kinpara nagahashika* mosquitoes on the *mizuhiki* flowers.
	2	57	そして、詳しく観察しました。
			And (I) observed (them) in detail.
28	1	58	その結果、ミズヒキの花は朝9時頃から12時にかけて満開になることがわかりました。
			The results were that *mizuhiki* flowers come into full bloom in the morning from 9:00 through 12:00.
G	S	T	Modified elaborated
	2	59	そして、それに合わせてキンパラナガハシカの雄が午前中だけ**吸蜜**(3)に来る事が分かりました。

			And it was found that in response to this, male *kinpara nagahashi ka* mosquitoes came for **feeding on nectar** (3) only in the morning.
29	1	60	富山ではいつも午前中に観察していました。
			I had always observed in the morning in Toyama.
	2	61	金沢の**里山**(3)では昨年以前は午後に観察していました。
			In Kanazawa, I had observed in the afternoon in the **woodlands** (3) of Kanazawa before last year.
	3	62	そんな単純なことが結果を左右していました。
			Such a simple thing affected the results.
	4	63	言い換えると、金沢では午前に観察していなかったから、見つけられなかったということです。
			In other words, the reason I did not find them in Kanazawa was because I was not observing in the morning in Kanazawa.
30	1	64	ミズヒキとは対照的にヒヨドリバナには午後2時ごろから5時ごろに蚊が**吸蜜**(3)に来ます。
			In contrast to *mizuhiki*, mosquitoes come to *hiyodori bana* flowers for **feeding on nectar** (3) between 2:00 and 5:00 in the afternoon.
31	1	65	このことから、自然の中で季節ごとに開花が変わりますが、カはそれに合わせて、利用する花を変えることがわかりました。
			From this, it was found that in nature full bloom periods change every season, and that mosquitoes adjust accordingly to this to change the flowers that they use.
	2	66	また一日の中でも時間帯によって利用する花を変えることが明らかになりました。
			Also, it became clear that they (mosquitoes) change the flowers that they use depending on the time of day.
	3	67	言い換えると、カは季節や時間で蜜を吸う花を変えているということです。
			In other words, mosquitoes change the flowers they suck nectar from based on season and time.
32	1	68	人にとってカは病気を媒介する害虫として有名です。
G	S	T	Modified elaborated
			For humans, mosquitoes are notorious as pests that transmit diseases.
	2	69	でも、実は、カはこれらの**野草**(3)の花粉媒介者、花粉を運ぶものでもあります。

			But, in fact, the mosquitoes are pollinators of this **wild grass** (3), or carriers of the plant pollen.
	3	70	血を吸わない雄が植物の花粉を媒介すると聞いて驚く人もいるかもしれません。
			There may be people who are surprised to hear that males, which do not suck blood, transfer the plant pollen.
33	1	71	また、将来**樹液**(3)を吸う蚊も観察できるかもしれません。
			Also, we may be able to observe mosquitoes that suck **tree sap** (3) in the future.
34	1	72	カの研究の中では、カ類は、世界で 3500 種いると言われています。
			In mosquito research, it is said that there are 3500 types of mosquito species in the world.
	2	73	しかし、このように、生態系の中での役割の理解はまだまだ進んでいません。
			However, as mentioned here, understanding of the roles in the ecological system of the mosquito is not yet advanced.
	3	74	これから面白い発見が色々出て来ると思います。
			I think that a variety of interesting findings will continue to emerge from now on.

Appendix D. Comprehension questions

R: Replication item; S: Synthesis item; I: Inferential item

Talk 1

Instructions 今、聞いた話を基に次の質問に答えてください。

Please respond to the following questions based on the talk that you just listened to.

答えは一つだけ選んでください。

Please select only one choice for your answer.

Sec1 1 R 国際開発学では、どんなことを勉強しますか。

In international development studies, what kind of things do they study?

a アメリカの歴史 History in the U.S.

b 貧しい国の開発のやり方

How to develop poor countries.

c タイの仕事

Work in Thailand.

d 心臓病やガンの治療

Treatment for heart diseases and cancers.

2 R 次の４つのうち、国際開発学の分野に入るものは何ですか。

Of the following four choices, what is included in the field of international development studies?

a バンコクオフィスの問題解決

Conflict resolution at the office in Bangkok.

b アメリカの大学の問題解決

Problem solving at universities in the U.S.

c 文化の違う国を理解する援助

Assistance understanding countries with different cultures

d 食べ物のない国への援助

Aid for countries without food

Sec 2 3 R この人が海外ボランティアしたところの特徴を一つ選んでください。

Please choose one characteristic of the places where this person did overseas volunteer work.

a　国際開発学の学生がたくさんいるところでした。
A place where many students in international development studies were

b　木がたくさんあるところでした。
A place where there are many trees.

c　水や電気が不足しているところでした。
A place where water and electricity were scarce.

d　いい大学があるところでした。
A place where good universities were found.

4　S　この人の一番興味がある国は、どれですか。
In which type of country is this person most interested?

a　貧しくて、危ない国。
Poor and dangerous countries.

b　貧しくても、仕事がある国。
Countries where jobs are found even if they are poor.

c　日本と貿易をしている国。
Countries which have trading relationships with Japan.

d　もともとイギリスの土地だった国。
Countries which were originally part of the U.K.

Sec 3　5　S　バンコクオフィスの仕事は、どうしてよかったですか。
Why was the job at the office in Bangkok good?

a　いろいろな教科書を見たからです。
Because he saw various textbooks.

b　歴史の長い国際機関だったからです。
Because it was an international organization which has a long history.

c　自分の知識と関心が広がったからです。
Because it expanded his knowledge and interests.

d　タイの仕事がいろいろできたからです。
Because he could do various work in Thailand.

6　S　この話のタイトルは、次のうち、どれが一番いいでしょうか。
Which of the following is the most appropriate title for this lecture?

a　国際開発と海外ボランティア

International development and overseas volunteers

b 国際開発学と将来の仕事

International development studies and careers in the future

c 国際開発と労働の問題

International development and labor problems

d 国際開発学とバンコクオフィス

International development studies and the office in Bangkok

7 I 国際開発学を学ぶ学生がした方が一番いいものは、次のうちどれでしょうか。

Which of the following is the best thing that students studying international development studies to do?

a 途上国での海外ボランティア

Volunteer overseas in developing countries.

b 工場での仕事

Work at a factory.

c 子どもの世話

Do childcare.

d ヨーロッパ旅行

Take a trip to Europe.

8 I 国際労働機関が行う仕事は、次のうちどれでしょうか。

Which of the following is work that involves ILO?

a 母親と子供を守る環境作り

Creating an environment to protect mothers and children.

b 世界と地域の労働についてまとめる仕事

Working to organize labor between global and local areas.

c 地域のための病院作り

Creating a hospital for a local area.

d きれいな水のための井戸掘り

Digging a well for clean water.

9 I この人の話によると、次のどの文が正しいですか。

According to this person's lecture, which of the following sentences is correct?

a 途上国は貧しいので、留学経験をもった優秀な人はいません。
Because developing countries are poor, there is no talented person who has experience with study abroad.

b 先進国と途上国の貿易ビジネスは、途上国の人たちのためになっています。
Trade business between developed and developing countries are beneficial for people in developing countries.

c 日本のような先進国は、国際労働機関には、関係がありません。
Developed countries, such as Japan, have nothing to do with ILO.

d 途上国と先進国の違いは重要な問題です。
The difference between developing and developed countries is an important problem.

Talk 2

Sec1 1 R 血を吸う虫は人間にとって どうして重要だと言っていますか。
Why did he say that blood sucking insects are important for humans?

a 雌と雄で血の吸い方が違うからです。
Because the way of sucking blood is different between females and males.

b 近くにたくさんいるからです。
Because there are many nearby.

c これらの虫が人間の栄養になるからです。
Because these insects become nutrition for humans.

d 人間が病気になるからです。
Because humans become sick.

2 R 蚊の雌は何のために血を吸いますか。
For what purpose do female mosquitoes suck blood?

a 雄のためです。
For males.

b 卵のためです。
For eggs.

c 自由な生活のためです。
For free lives.
d 健康のためです。
For health.

Sec 2 3 R 暑い国の病気は、次のうち、どれですか。
Which of the following is a disease in hot countries?
a フィラリア
filaria (heartworm)
b デング熱
dengue fever
c 心臓病
heart disease
d 日本脳炎
Japanese encephalitis

4 S 病気を移す蚊について、正しい文を一つ選んでください。
Please choose one correct sentence about mosquitoes that transmit diseases.
a 全部の蚊は、人に病気を移します。
All mosquitoes transmit diseases to humans.
b 蚊は動物の病気を人に移しません。
Mosquitoes do not transmit animal diseases to humans.
c 魚の血を吸う蚊は、魚の病気を人に移します。
Mosquitoes that suck fish blood transmit fish diseases to humans.
d 病気を移す蚊の研究は多いです。
There are many studies about mosquitoes that transmit diseases.

Sec 3 5 S 蚊の雌と雄について、正しい文を一つ選んでください。
Please choose one correct sentence about mosquitoes that transmit diseases.
a 雌の蚊は、花粉を運びます。
Female mosquitoes carry pollen.
b 雄の蚊は、花粉を運びます。
Male mosquitoes carry pollen.
c 雄の蚊は、血を吸います。

Male mosquitoes suck blood.

d 蚊の研究は、雌、雄に関係なく、多いです。

There are many studies about mosquitoes regardless of whether they are females or males.

6 S この話のタイトルは、次のうち、どれが一番いいでしょうか。

Which of the following is the most appropriate title of this lecture?

a 蚊と動物の関係。

The relationship between mosquitoes and animals.

b 蚊と病気の関係

The relationship between mosquitoes and diseases.

c 血を吸う蚊と吸わない蚊。

Mosquitoes that suck blood and mosquitoes that do not.

d 蚊と植物の関係。

The relationship between mosquitoes and plants.

7 I 蚊の雄を野外で研究するのは、どうして難しいですか。

Why is it difficult to research male mosquitoes out fields?

a 雄は、雌より少ないからです。

Because there are fewer males than females.

b 雄は動物の近くに住んでいるからです。

Because mosquitoes are living near animals.

c 雄がどこで見つけられるのかがよくわからないからです。

Because we do not know where we can frequently find males.

d 雄は魚の血を吸うからです。

Because male mosquitos suck blood of fish.

8 I 雄の蚊は、何にとって大切ですか。

For what are male mosquitoes important?

a 雄の蚊は、花にとって大切です。

Male mosquitoes are important for flowers.

b 雄の蚊は、木にとって大切です。

Male mosquitoes are important for trees.

c 雄の蚊は、山にとって大切です。

Male mosquitoes are important for mountains.

d 雄の蚊は、動物にとって大切です。

Male mosquitoes are important for animals.

9 I この人の話によると、次のどの文が正しいですか。
According to this person's lecture, which of the following sentences is correct?

a 蚊が移す病気を心配しなくてもいいとわかりました。
We found that there is no need to worry about diseases transmitted by mosquitoes.

b 蚊の研究は、水の近くがよさそうです。
It appears better to conduct mosquito studies near water.

c 蚊の研究は、建物の中がよさそうです。
It appears better to conduct mosquito studies in buildings.

d 世界には、ほかにも変わった蚊がいそうです。
It appears that there could be other strange mosquitoes in the world.

Talk 3

Sec1 1 R 宇宙物理学では、何を研究しますか。
In astrophysics, what do they study?

a 宇宙の星やそれに関係するいろいろなことを研究します。
They study stars in outer space and various things related to them.

b 宇宙の星と私たちの生活を研究します。
They study stars in outer space and our lives.

c 宇宙の星を星空から研究します。
They study stars in outer space under a starlit sky.

d 宇宙の星と人工衛星を研究します。
They study stars in outer space and satellites

2 R ガンマ線で、何ができますか。
What can they do with gamma rays?

a 携帯電話がかけられます。
They can make a cell phone call.

b ブラックホールが調べられます。
They can investigate a blackhole.

c 人工衛星が見られます。
They can see a satellite.
d ヒーターとして使えます。
They can use there as a heater.

Sec2 3 R 雷が大気中で起きる電気現象だといつわかりましたか。
When did they discover that lightning is an electric phenomenon happening in the air?
a 最近わかりました。
They discovered it recently.
b 数十年前にわかりました。
They discovered it a few decades ago.
c 数百年前にわかりました。
They discovered it a few hundred years ago.
d まだわかっていません。
They have not discovered it yet.

4 S この人は雷を調べることについて、どう考えていますか。
What does this person think about studying lightning?
a 静電気なので、やさしいと考えています。
He thinks it is easy because it is static electricity.
b 目に見える光なので、やさしいと考えています。
He thinks it is easy because of visible light.
c ブラックホールと同じで、難しいと考えています。
He thinks it is as difficult as studying a blackhole.
d 難しいかもしれないけれど、ぜひ調べてみたいと考えています。
He thinks that it may be difficult, but he is really willing to study it.

Sec 3 5 S 雷と宇宙には、どのような関係があるかもしれませんか。
What kind of relationship might there be between lightning and outer space?
a ブラックホールと雷に関係があるかもしれません。
There may be a relationship between a blackhole and lightning.
b 宇宙線シャワーと雷に関係があるかもしれません。

There may be a relationship between cosmic radiation and lightning.

c　宇宙の強い静電気の空間と雷に関係があるかもしれません。

There may be a relationship between strong static electricity in outer space and lightning.

d　宇宙ガンマ線と雷に関係があるかもしれません。

There may be a relationship between gamma rays in outer space and lightning.

6　S　この話のタイトルは、次のうち、どれが一番いいでしょうか。

Which of the following is the most appropriate title for this lecture?

a　空港の検査と雷の関係

Cosmic radiation and lightning

b　雷と光の種類

Lightning and types of light

c　静電気と雷

Static electricity and lightning

d　宇宙を見るガンマ線と雷

Techniques to investigate outer space and lightning

7　I　この人の話から、光についてわかることは何ですか。

According to this person's lecture, what is known about light?

a　目に見えない光は、私たちの生活の中で、結構使われています。

We use invisible light in both our lives.

b　人工衛星で目に見える星を調べています。

They are investigating visible stars with satellites.

c　光のエネルギーは、みんな同じです。

Energy from light is all the same.

d　目に見える光が私たちの生活に一番身近です。

Visible light is the closest to our lives.

8　I　この人の話から、雷についてわかることは何ですか。

According to this person's lecture, what is known about lightning?

a　雷は雨の時に水を通る電気現象かもしれません。

Lightning may be an electronic phenomenon that is transmitted through water during rain.

b 雷は空で起こるので、地上で起こる静電気の現象とは基本的に違います。

Because lightning occurs in the sky, it is fundamentally different from the static electricity phenomenon that occurs on the ground.

c 雷がどのように起こるかは長い間研究されているのに、まだわかっていません。

Although how lightning occurs has been studied for a long time, it has not yet revealed.

d 雷を直接調べて、電流が流れて、危ない事故がありました。

There was a dangerous accident with streaming electric current when lightning was directly investigated.

9 I この人の話によると、次のどの文が正しいですか。

According to this person's lecture, which of the following sentences is correct?

a この人のプロジェクトは、人工衛星も使います。

This person’s project uses a satellite as well.

b この人のプロジェクトは、新しい方法で地球のことを調べます。

This person’s project studies the Earth through a new method.

c この人のプロジェクトでは、昔からあるやり方で地球のことを調べます。

This person’s project studies the Earth using traditional methods that have existed for a long time.

d この人のプロジェクトは、静電気も使います。

This person’s project uses static electricity.

Talk 4

Sec1 1 R ワシントン日本語継承センターはアメリカのどこにありますか。

Where in the U.S. is the Japanese heritage center located?

a ワシントン州

State of Washington

b　メリーランド州

State of Maryland

c　ニューヨーク州

State of New York

d　アメリカの私立学校で教えています。

He is teaching at a private school in the U.S.

2　R　継承センターについて正しい文を一つ選んでください。

Please choose one correct sentence about the heritage center.

a　授業は毎週土曜日と日曜日の午前中にあります。

They have classes in the morning every Saturday and Sunday.

b　校舎は、センターの建物です。

The school building is the center's building.

c　センターの子どもたちの数は、始まった時よ増えました。

The number of children at the center has increased since its beginning.

d　クラスは、子供のためだけです。

Their classes are only for children.

Sec 2　3　R　継承センターの子どもがいえでする宿題に入っていたものを一つ選んでください。

Please choose one answer that was included in homework that children at the center do at home.

a　聞き取り練習。

Listening practice.

b　漢字練習。

Kanji practice.

c　歌の練習。

Singing practice.

d　料理の練習。

Cooking practice.

4　S　日本語継承センターと日本語補修学校は、何が違いますか。

What is a difference between the Japanese Heritage Center and Japanese Language School?

a　子どもの年齢が違います。

Children's ages are different.

b 使う言語が違います。
The language used is different.

c 学校のある地域が違います。
The area where the school is located is different.

d 教育の内容とやり方が違います
The content and way of education are different.

Sec 3 5 S 継承センターの問題は、どれですか。
Which of the following is a problem for the Heritage Center?

a センターの校舎がないことです。
The fact that the center does not have a school building.

b 日本語補習校があることです。
The fact that there is Japanese Language School.

c 教科書などがなく、手作りになることです。
The fact that they have no textbooks and all materials are handmade.

d 教師の仕事の量が少ないことです。
The fact that the work load of teachers is light.

6 S この話のタイトルは、次のうち、どれが一番いいでしょうか。
Which of the following is the most appropriate title for this lecture?

a 日本語継承センターの特色と問題点
Characteristics and problems of the Japanese Heritage Center.

b 日本語継承センターの歴史
History of the Japanese Heritage Center.

c 日本語継承センターの教育
Education in the Japanese Heritage Center.

d 日本語継承センターの問題点
Problems at the Japanese Heritage Center.

7 I 継承センターと比べて、日本語補習学校はどんな学校のようですか。
Compared to the Heritage Center, what is Japanese Language School like?

a 少人数クラスです。

Small classes.

b　自主的活動が多いです。

Many voluntary activities.

c　日本の学校とほとんど同じです。

Almost the same as schools in Japan.

d　宿題はほとんどありません。

Almost no homework.

8　I　次の4人の子どものうち、どの子供が一番継承センターに合っていると思いますか。

Which of the following children do you think would best fit at the Heritage Center?

a　土曜日と日曜日にサッカーをする子ども

Children who play soccer on Saturdays and Sundays.

b　積極的にいろいろなことに挑戦したい子ども。

Children who want to challenge themselves actively with various things.

c　日本の大学を受験したい子ども。

Children who want to take an entrance exam for a Japanese university.

d　日本のアニメを勉強したい子ども。

Children who want to study Japanese animation.

9　I　この人の話によると、次のどの文が正しいですか。

According to this person's lecture, which of the following sentences is correct?

a　継承センターは、人気が出ているようです。

It seems that the Heritage Center is getting popular.

b　継承センターで、子どもが友達を作るのは難しそうです。

It seems that it is hard for children to make friends at the Heritage Center.

c　継承センターは、文部科学省の教育内容にも合わせようとしています。

The Heritage Center is trying to meet the education content requirements of the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

d 継承センターは、宿題が一番大切だと考えているようです。

It seems that the Heritage Center thinks that homework is most important.

Appendix E. Form-recognition test: Choices

			a		b		c		d	
1	シンクタンク	Think-tank	sinktank	シンクタンク	sunktank	スンクタンク	senktank	センクタンク	sonktank	ソンクタンク
2	発祥(はっしょう)	beginning	tasshoo	たっしょう	kasshoo	かっしょう	sasshoo	さっしょう	hasshoo	はっしょう
3	民間(みんかん)セクター	private sector	munkan sekutaa	むんかんセクター	minkan sekutaa	みんかんセクター	monkan sekutaa	もんかんセクター	menkan sekutaa	めんかんセクター
4	インターン	intern	intaan	インターン	antaan	アンターン	untaan	ウンターン	entaan	エンターン
5	受(う)け皿(ざら)	receiver,	ukuzara	うくざら	ukizara	うきざら	ukezara	うけざら	ukazara	うかざら
6	インフラ	infrastructure	onfura	オンフラ	infura	インフラ	enfura	エンフラ	unfura	ウンフラ
7	NGO	NGO	enujiioo	エヌジーオー	emujiioo	エムジーオー	inujiioo	イヌジーオー	imujiioo	イムジーオー
8	是正(ぜせい)	correction	sasei	させい	sesei	せせい	zasei	ざせい	zesei	ぜせい
9	哺乳類(ほにゅうるい)	mammal	honyoorui	ほにょうるい	honyaarui	ほにゃあるい	honyuurui	ほにゅうるい	honoorui	ほのうるい
10	成虫(せいちゅう)	imago	seechuu	せいちゅう	teechuu	ていちゅう	keechuu	けいちゅう	reechuu	れいちゅう
11	ボウフラ	wiggler	buufura	ブウフラ	boofura	ボウフラ	guufura	グウフラ	goofura	ゴウフラ

12	棲息（せいそく）	inhabitation	seeshiku	せいしく	seeseku	せいせく	seesuku	せいすく	seesoku	せいそく
13	樹液（じゅえき）	tree sap	shueki	しゅえき	shoeki	しょえき	joeki	じょえき	jueki	じゅえき
14	里山（さとやま）	woodland near town	satsuyama	さつやま	satayama	さたやま	satoyama	さとやま	sateyama	さてやま
15	吸蜜（きゅうみつ）	nectar sucking	kyoomitsu	きょうみつ	kyuumitsu	きゅうみつ	kyaamitsu	きゃあみつ	kuumitsu	くうみつ
16	野草（やそう）	wild grass	yasoo	やそう	rasoo	らそう	nasoo	なそう	masoo	まそう
17	望遠鏡（ぼうえんきょう）	telescope	buuenkyoo	ぶうえんきょう	beeenkyoo	べいえんきょう	booenkyoo	ぼうえんきょう	baaenkyoo	ばあえんきょう
18	可視光（かしこう）	visible light	kushikoo	くしこう	keshikoo	けしこう	koshikoo	こしこう	kashikoo	かしこう
19	銀河系（ぎんがけい）	the Galaxy	kingakee	きんがけい	gingakee	ぎんがけい	gungakee	ぐんがけい	jingakee	じんがけい
20	太古（たいこ）	ancient times	taiko	たいこ	teiko	ていこ	toiko	といこ	tsuiko	ついこ
21	エニグマ	enigma	amiguma	アミグマ	aniguma	アニグマ	eniguma	エニグマ	emiguma	エミグマ
22	放電（ほうでん）	electric discharge	huuden	ふうでん	hooten	ほうてん	huuten	ふうてん	hooden	ほうでん
23	バルーン	balloon	boruun	ボルーン	baruun	バルーン	beruun	ベルーン	buruun	ブルーン
24	粒子（りゅうし）	particle	ryooshi	りょうし	ruushi	るうし	ryuushi	りゅうし	rooshi	ろうし
25	黒板（こくばん）	blackboard	kokoban	ここばん	kokaban	こかばん	kokiban	こきばん	kokuban	こくばん

26	学(まな)び舎(や)	school	manabiya	まなびや	mamabiya	ままびや	manebiya	まねびや	manibiya	まにびや
27	下校(げこう)	going home from school	gekoo	げこう	gukoo	ぐこう	kekoo	けこう	gikoo	ぎこう
28	ロゴ	logo	rugo	ルゴ	rago	ラゴ	rogo	ロゴ	rigo	リゴ
29	抜粋(ばっすい)	excerpt	bossui	ぼっすい	bassui	ばっすい	bessui	べっすい	bussui	ぶっすい
30	ドリル	drill notebook	doiru	ドイル	dariru	ダリル	doriru	ドリル	dairu	ダイル
31	縦割(たてわ)り	classes with mixed age	takewari	たけわり	tachiwari	たちわり	tatsuwari	たつわり	tatewari	たてわり
32	推奨(すいしょう)	recommendation, endorsement	shiishoo	しいしょう	suishoo	すいしょう	jiishoo	じいしょう	zuishoo	ずいしょう

Appendix F. Meaning recognition sentence test

Talk 1

1 シンクタンク think-tank

a バンコクのオフィスは、パリのオフィスとシンクタンクをして、情報を交換していました。

The office in Bangkok was exchanging information with the office in Paris using a think-tank.

b 太平洋では、第一次世界大戦からのシンクタンクがまだ見つかるそうです。

We heard that think-tanks from WWI are still discovered in the Pacific Ocean.

c タイのシンクタンクに、珍しいウミガメが保護されているそうです。

We heard that rare sea turtles are protected at think tanks in Thailand.

d 日本と中国の経済関係について、有名なシンクタンクが国際会議を開いたそうです。

We heard that a famous think-tank hosted an international conference about the economic relationship between Japan and China.

2 発祥　はっしょう beginning

a 国際労働機関の発祥は、お金のない地域の人たちの問題の解決です。

The origin of ILO is to solve issues for people in poor areas.

b 日本の小学校の発祥の地は、京都です。

The original site of Japanese elementary schools is Kyoto.

c 学生の時に海外で経験したことは、将来の仕事の発祥になります。

Things you experienced overseas in your university days will become the origins for your future career.

d 国際開発学の発祥では、世界の経済的な格差について、取り組みます。

In the origin of international development studies, we focus on global economic disparity.

3 民間セクター private sector

a 日本でホテルが高すぎる場合は、民間セクターで宿泊が可能です。

If the cost of hotels is too high in Japan, it is possible for you to stay in the private sector.

b 民間セクターを開発するのは、商社や銀行の仕事です。

It is the job of trading companies and banks to develop the private sector.

c 太平洋の島国をグループに分けるときに民間セクターを利用すると便利です。

It is convenient for you use the private sector when you divide islands in the Pacific Ocean into groups.

d 新しい工場は、民間セクターの新しいプロジェクトのために建てられています。

A new factory is built for a new project in the private sector.

4 インターン intern

a 国際開発学を勉強する学生は、インターンをビデオで見て、就職先を探します。
Students who study international development watch intern in videos and look for work places.

b 貧しい地域では、インターンのあるうちが少ないので、不便です。
In poor areas, it is inconvenient because there are fewer houses with interns.

c アメリカでは、夏休みに、政府や会社でインターンの仕事をする学生が多いです。
In the U.S., there are many students who work as interns at offices in the governments and companies during their summer breaks.

d 職場で労働環境を整える時に、少し休んだりできるインターンを作る会社が増えています。
When adjusting the labor environment at a workplace, companies that create an intern where employees can take a little break are increasing.

5 受け皿 うけざら receiver

a 海外ボランティアが来たときは、地元の町全体が受け皿になったそうです。
We heard that the entire local community became a host for volunteers from overseas when they came.

b イギリスは元々多くの植民地を抱えていましたから、お茶を植民地に輸出する受け皿でした。
Britain was a host that exported tea to colonies, because it originally had many colonies.

c 東南アジアでは、学生は、受け皿を使って、就職先を探していました。
In Southeast Asia, students have looked for work using a host.

d タイの会社は、従業員を小さい工場に送る受け皿をしていました。
A company in Thailand took responsibility for sending employees to small factories.

6 インフラ infrastructure

a インフラが賑やかかどうかは、地域の人たちの雇用には全く関係がありません。
It has nothing to do with employment of local people whether or not infrastructure is lively.

b ケニアでは、インフラの中から建設機械を出して、道路を作ります。
In Kenya, they create roads by taking out construction equipment from an infrastructure.

c 貧しい国では、人々が病気にならないようにインフラを飲みます。
In poor areas, people drink infrastructures so that they do not become sick.

d 太平洋の小さい島国は、観光ビジネスで得る利益を国のインフラのために使うべきです。

Small island countries in the Pacific Ocean should use profit obtained from the tourist industry for infrastructure in the countries.

7 NGO

a 国によっては、T シャツや短いスカートで学校に行くのは NGO のところもあります。

Depending on the country, there is an NGO place where one goes to school wearing T-shirts and short skirts.

b 食料があまりない貧しい地域では、NGO を使って、料理の味をよくします。

In the poor areas where there is not much food, the taste of food is made better by using NGO.

c 代表的な NGO は、環境保護や紛争で家族を亡くした人の保護、自然保護などを行っています。

Typical NGOs are doing environmental protection, protection for people who lost family members due to conflicts, and nature conservation.

d 日本の NGO は、期限の切れた日本人のパスポートを新しいものにすることができます。

Japanese NGOs can make expired Japanese passports new.

8 是正　ぜせい correction

a オリンピックのための重要課題の一つは、外国人の入国システムの是正です。

One of the important issues for the Olympics is the correction of the immigration control system for foreign visitors.

b 日本では引きこもりと呼ばれる大人が増えていますから、是正が必要です。

Correction is required because adults called personal withdrawals from the society have been increasing in Japan.

c 海外ボランティアのグループは、木を植えて、その町の公園の是正を助けました。

A volunteer group from overseas planted trees and helped with a correction of the town's park.

d 私は商社で、日本の建設機械を輸出する仕事の是正を行って、利益を増加させました。

At a Japanese trading company, I worked on corrections for the job of exporting Japanese construction equipment and increased profits.

Talk 2

9 ほにゅうるい mammal

a 子どもがよく食べるものは、哺乳類です。

Things that children often eat are mammals.

b	人は、哺乳類です。
Humans are mammals.
c	哺乳類が動物に病気を移します。
Mammals transmit diseases to animals.
d	1970 年代には、ウィルスの病気の患者は、哺乳類を使いました。
In the 1970s, patients sick with viruses used mammals.

10	せいちゅう imago
a	動物の子どもは、成虫になると、自由生活をします。
Young animals live freely once they become imagoes.
b	蚊の雄は、花の成虫から蜜を取ります。
Female mosquitoes take nectars from imagoes of flowers.
c	蚊は成虫になると、水中から出て来ます。
Mosquitoes come out of water once they become imagoes.
d	蚊は、成虫といっしょにいる動物に病気を移します。
Mosquitoes transmit diseases to animals that are with imagoes.

11	ボウフラ wiggler
a	日本には、花粉を運ぶボウフラがいます。
In Japan, there are larvae that carry pollen.
b	ボウフラは、暑い国の病気です。
Larvae are a disease in hot countries.
c	ボウフラは、午後に花が咲いて、満開になります。
Larvae bloom in the afternoon and go into full bloom.
d	魚は、ボウフラを食べます。
Fish eat larvae.

12	棲息　せいそく inhabitation, living
a	キンパラナガハシカの棲息が金沢でも確認されました。
Existence of *kinpara nagahashika* was confirmed in Kanazawa as well.
b	1970 年代には、暑い国の病気の棲息が日本でも確認されました。
In the 1970's, the existence of diseases from hot countries was also confirmed in Japan.
c	蚊の行動は、かなり詳しく棲息をされてきました。
Behaviors of mosquitoes come from fairly detailed habitat.
d	蚊がどの動物の血を吸うかは、棲息によって決まっています。
Which animals' blood that mosquitoes suck is determined by their habitat.

13　樹液　じゅえき tree sap

a　樹液の患者が、日本でも報告されています。
Patients with tree sap have been reported in Japan as well.

b　魚の樹液を吸う蚊もいます。
There are mosquitoes that suck the (tree)sap of fish.

c　蚊は樹液を吸って、生きることもあるそうです。
We heard that there are cases when mosquitoes live by sucking tree sap.

d　雌も雄も血を吸う虫は、樹液の中にいます。
Both female and male mosquitoes that suck blood are in tree sap.

14　さとやま woodlands near town

a　金沢では、里山で動物をよく見ました。
In Kanazawa, I often saw animals at the woodlands.

b　日本にも、卵ができない里山がありました。
In Japan also, there are woodlands that cannot produce eggs.

c　雌の蚊は、里山を通して、病気を移します。
Female mosquitoes transmit diseases through woodlands.

d　シラミやノミのような里山は、あまり研究されていません。
Woodlands, such as lice and fleas, have not been studied much.

15　吸蜜　きゅうみつ　nectar sucking

a　動物の子どもは、吸蜜を通して、生きている環境がわかります。
Living environment of young animals is revealed through nectar sucking.

b　ウィルスの病気の代表的なものに、吸蜜があります。
A typical viral disease is nectar sucking.

c　吸蜜は、私たち人間にとって大変重要です。
Nectar sucking is very important for us humans.

d　血を吸わない蚊は花が咲く時間に吸蜜に来ます。
Mosquitoes that do not suck blood come for nectar sucking at the time when flowers bloom.

16　野草　やそう　wild grass

a　蚊は、野草に合わせて、血を吸います。
Mosquitoes suck blood in coordination with wild grass.

b　動物や虫の観察は、野草があるところでしています。
I have been conducting observations of animals and insects at a place where wild grass is found.

c　魚を研究すると、野草の病気がわかります。

When you study fish, you understand diseases of wild grass.

d 他の種類の子どもを食べる野草もあります。

There is also wild grass that eats the youth of other kinds.

Talk 3

17 望遠鏡 ぼうえんきょう telescope

a この望遠鏡からやや下の方を見ると、野生の動物や鳥がよく見えます。

You see wild animals and birds when looking down a little from this telescope.

b 望遠鏡の鏡に、雲を写して、正確に雲を観測することができます。

You can accurately observe cloud by capturing cloud (image) on the mirror of the telescope.

c 自分の目で星を観測するには、望遠鏡を使った方がよく見えます。

In order to observe stars with our own eyes, using the telescope is better to see them clearly.

d 望遠鏡の前に立って、出かける前の服装の確認をしています。

I am checking my clothes before going out by standing in front of the telescope.

18 可視光 かしこう visible light

a 雲の中から見える光は、可視光ではありません。

The light that can be seen in the clouds is not visible light.

b 可視光があるので、静電気の空間が強くても、イタッとなりません。

Due to visible light, you do not feel pain even if space with static electricity is strong.

c 空港の荷物検査で使われる光は、可視光ではありません。

The light used for the baggage inspection at the airport is not visible light.

d 可視光を使って、ブラックホールが観測できます。

We can observe a blackhole by using visible light.

19 銀河系 ぎんがけい the galaxy

a 銀河系を観測する時、静電気の空間があると、観測しやすいです。

When observing the Galaxy, it is easier to observe having space with static electricity.

b 銀河系には、太陽のような星がたくさんあると言われています。

It is said that there are many stars like sun in the Galaxy.

c 銀河系の人たちが、宇宙線シャワーと雷に関係があると考えています。

People in the Galaxy think that there is a relationship between cosmic radiation and lightening.

d 銀河系は、エネルギーの高い光に関係があるので、携帯電話にも関係があります。

Because the Galaxy has a relationship with high energy light, it also has a relationship with cell phones.

20 太古　たいこ　　ancient times

a 雲の中の強い静電気の空間は、太古と思われる状態と関係があります。

Space with strong static electricity in the clouds has a relationship with a state that is thought to be ancient times.

b 最近の研究で、ガンマ線と太古の関係が明らかになりました。

Recent studies revealed a relationship between gamma ray and ancient times.

c 地上からのブラックホールの観測は、地球の太古が厚いので、難しいです。

An observation of a blackhole from the ground on earth is difficult due to thick ancient times on the Earth.

d 宇宙について、太古に生きた人たちも様々な方法を使って説明しようとしました。

Regarding the outer space, people who lived in ancient times also tried to explain it using various ways.

21 エニグマ enigma

a 高エネルギーの宇宙線シャワーは、エニグマによって速められます。

Cosmic radiation with high energy is accelerated by an enigma.

b 宇宙には、まだ研究で明らかにされていないエニグマがたくさんあります。

In the outer space, there are many enigmas that have not been revealed in studies yet.

c 人工衛星にエニグマの検出器を載せて、実際の観測は地球の外で行います。

Loading an enigma detector on a satellite, an actual observation is conducted outside the Earth.

d 静電気の空間がない時に、雷はあっても、エニグマはありません。

When there is no space with static electricity, there is no enigma while there is lightening.

22 放電　ほうでん electric discharge

a 冬にホテルのドアなどに触ると、空気中に電気が流れて、放電が起きます。

When touching a thing, such as a door in a hotel, an electric discharge occurs streaming electricity in the air.

b コンピュータを長く使っていると、放電が起きます。

When using a computer for a long time, an electric discharge occurs.

c ヒーターから赤外線が出ると、放電が起きて暖かくなります。

When infrared radiation emits from a heater, it becomes warmer occurring an electric discharge.

d 携帯電話に使われている電波の関係で、雨の時に放電が起こることもあります。

With the relation of an electric wave used for cell phones, there is a case where an electric discharge occurs when it rains.

23　バルーン balloon

a　静電気の計測を行うのに、バルーンというガンマ線の一種を使うそうです。

We heard that a type of gamma rays called a balloon is used to conduct a measurement of static electricity.

b　バルーンは、ブラックホールや中性子星のような特殊な星まで調べられます。

A balloon can investigate special stars, such as blackholes and neutral stars.

c　バルーンには、人が数人乗れますが、事故で落ちると、大変なことになります。

A few people can ride in the balloon, but it becomes a disaster if it falls as an accident.

d　バルーンで、エネルギーの高い光も低い光も、簡単に観測できます。

We can observe both high and low energy light with a balloon.

24　粒子　りゅうし particle

a　粒子のない世界は、水などの液体の世界です。

The world without particles is the world of liquid, such as water.

b　静電気でパチパチ、イタッとなるのは、目で見られるほどの粒子が体にあたるからです。

The reason why we have the crackling and the pain of static electricity is because particles visible hit the body.

c　粒子があれば、様々なエネルギーの光は、必要ないかもしれません。

If there are particles, light for various energy may not be needed.

d　宇宙には粒子があって、これが環境による条件で、速められてシャワーになります。

In the outer space, there are particles, and they are accelerated to become shower with environmental conditions.

Talk 4

25　黒板　こくばん blackboard

a　子どもたちは、黒板に書かれた文を自分のノートに書きました。

Children wrote the sentences written on the blackboard in their notebooks.

b　日本料理を作る時に、いい黒板を使うことは大切です。

When cooking Japanese cuisine, it is important to use a good blackboard.

c　ワシントンの桜まつりでは、子どもたちは黒板のステージで歌を歌いました。

At the cherry blossom festival in Washington DC, children sang songs on the blackboard stage.

d　少人数クラスでは、黒板を一人一人に見せることはできません。

In a small class, you cannot show the blackboard one-by-one.

26　学び舎　　まなびや school

a　学び舎で働いている大学生は、宿題ができない子どもたちを手伝います。
University students who are working at the school help children who cannot complete their homework.

b　学び舎で、子どもの教育に必要な便利なグッズが買えます。
At a school, you can buy convenient goods necessary for children's education.

c　漢字検定や日本語能力検定を作成しているのが、学び舎です。
It is a school that creates tests, such as kanji tests and Japanese proficiency test.

d　この学び舎は、50年もの間、子どもたちの勉強を見守りました。
This school oversaw children's study for 50 long years.

27　下校　げこう　　going home after classes

a　初めて入った小学校は難しかったので、少しやさしい小学校に下校しました。
Because the elementary school that I first entered was difficult, I transferred to a little easier elementary school.

b　家庭学習として、下校の問題集をする子どもたちもいます。
For home study, there are children who use study books they take home after school.

c　日本で子どもの下校の時は、たいてい友だちと一緒のことが多いです。
There are many cases where children are usually with their friends when they go home after school in Japan

d　継承センターの授業が始まってすぐに、下校の活動があります。
As soon as classes at the heritage center begin, they engage in an activity of going home after school.

28　ロゴ　logo

a　継承センターの催しは、いつもロゴで行われています。
Events at the heritage center always take place at the logo.

b　センターの子どもたちはロゴで東京タワーや東京駅のような有名な建物を作ります。
Children at the center create famous buildings, such as Tokyo Tower and Tokyo Station with logos.

c　センターの子どもは、ロゴを使って、漢字検定の準備をします。
Children at the center prepare for kanji tests using a logo.

d　ロゴには、その組織の理念やポリシーなども表されています。
A logo expresses organizational philosophy and policy as well.

29　抜粋　ばっすい excerpt

a　継承語教育では、抜粋のビデオで、練習することもあります。

In heritage language education, we sometimes practice with video excerpts.

b 継承センターの夏休みの宿題に、抜粋へ行くことがあります。

For homework during summer break at the heritage center, we sometimes go to excerpts.

c 検定教科書を使用するのに、子どもたちの保護者は抜粋に来て、子どもと話さなければなりません。

In order to use approved textbooks, children's' guardians must come to excerpts, and talk to children.

d 本の抜粋をしてもいいですが、本の名前は必ず一緒に書いておくべきです。

It is permitted to use excerpts from books, but you should write titles of the books together.

30 ドリル drill notebooks

a 家で勉強をする時は、ドリルで日本語を聞きます。

When studying at home, I listen to Japanese with drills.

b 子どもたちは、たいてい大きくて、明るい色のドリルが好きです。

Children usually like large and bright colored drills.

c 日本の学校では、様々なドリルを使って、子どもたちに宿題をさせます。

Japanese schools make children work on their homework using various drill notebooks.

d ドリルを聞いても、教科書に全く関係なく、子どもたちの勉強には役に立ちません。

Even if you listen to drills, they are useless for children's study because they have nothing to do with textbooks.

31 縦割り たてわり classes with mixed age

a クラスの中で縦割りにする時、背の高い子どもは前に座ります。

When holding classes with mixed age, tall children sit in front.

b 子どもが多い学校で、縦割りにするのは、難しそうです。

It seems difficult to adopt classes with mixed age at a school with many children.

c 日本の学校では、教室の中の机を縦割りにします。

In Japanese schools, desks in the classroom are divided vertically.

d 日本の食文化では、割りばしの縦割りを使った食べ方が少なくありません。

In Japanese food culture, there are not few ways to eat using a vertical split of disposal wooden chopsticks.

32 推奨 すいしょう recommendation endorsement

a 継承語センターを卒業する子どもたちは、日本とアメリカの文化をつなぐ推奨になると考えられます。

Children who graduate from the heritage center are thought to act as an endorsement that connects culture between Japan and the U.S.

b 継承語センターは、日本語をアメリカで勉強したいという推奨にこたえるために作られました。

The heritage center was created in response to a recommendation by those who wish to study Japanese in the U.S.

c 先生は、例文も多くて、漢字の説明もあるこの辞書が推奨だと言いました。

The teacher said that this dictionary that includes many examples sentences and explanation for kanji is the recommended one.

d アメリカでも日本文化のコンテストがたくさんあって、推奨の子どもが参加します。

There are many contests on Japanese culture in the U.S. as well, and the recommended children participate in them.

Appendix G. Meaning recognition definition test

Talk 1

1　シンクタンク　think-tank

a　珍しいウミガメが保護されているところ
a place where rare sea turtles are protected

b　国際的な組織の別々のオフィスでデータを同時に使えるようにすること
A system which makes it possible for different offices in an international organization to use data simultaneously

c　さまざまな分野の研究者が調査や分析を行っている組織のこと
An organization where researchers in various fields are conducting studies and analyses

d　第一次世界大戦に使われた海軍の船の一種
One type of naval ship that was used during WWI

2　発祥　はっしょう beginning

a　過去に経験したことが将来の役に立つこと
Things experienced in the past are useful for the future

b　何かが初めて出来たところのこと
The first place where something was established

c　ある研究分野の目標になるもの
A goal of a research field

d　ある研究分野の中心になるもの
The center of a research field

3　民間セクター private sector

a　国がしない仕事を国や会社などからお金をもらって行い、利益を求めない団体
A non-profit organization that does jobs that a government does not do by receiving money from the government and companies

b　銀行などがお金を出して作る組織で、利益も求めて、貧しい国で活動をする組織
An organization funded by companies, such as banks that also actively seek profits in poor countries

c　仕事のために、普通の家族のうちを安い値段で泊まらせてくれる宿泊制度
A lodging system which allows business travelers to stay in private family homes for a reasonable price

d　何かメンバーの多いグループを小さく分ける時に使われるコンピュータソフト
A computer software used for dividing many members in a group into small groups

4 インターン intern

a 家の前にある電話のようなもの
an item like a phone in front of a house

b 職場でお茶を飲んだりするための休める場所のこと
a place to rest for activities, such as drinking tea at the workplace

c 正式な従業員ではなくて、一時的に仕事のやり方を学ぶために入る実習生のこと
a trainee who is not an official employee and who temporarily joins to learn how to work.

d 国際開発学の学生のために作られた仕事の説明のこと
work instructions created for students who major in international development studies

5 受け皿 うけざら receiver

a 人やものを送った時に受け入れ先として責任を持ってくれるところのこと
organizations or people who take responsibility for accepting people and objects when they are sent

b 東南アジアで仕事を探す時の道具
A tool when you look for a job in Southeast Asia

c 会社が従業員を小さい工場に送る時の役割のこと
the role of a company when the company sends employees to small factories

d 植民地へ物を輸出する時、輸出の責任を取る国や人のこと
Parties, such as countries and people who take responsibility for exports when things are exported to colonies

6 インフラ infrastructure

a 貧しい国の人々が利用している栄養ドリンクで病気の予防になるもの
Nutrition drinks that are used by people in poor countries and that prevent diseases

b 住んでいる人の生活のために社会的に必要な公共の建物やシステムのこと
Public buildings and systems needed in society for residents' lives

c 建設機械を運ぶ時に入れるもののこと
A container where construction equipment is put where it is transported

d 町の中心でいろいろな店やレストランがあるところのこと
the center of a town where various shops and restaurants are located

7 NGO

a 政府の仕事を臨時ですることができる組織のこと
An organization that can temporarily do government work

b やってはいけないという意味の言葉
A word meaning you must not do

c 安い料理の調味料で、食料の無い貧しい地域でよく使われるもの

A cheap spice for cuisines that are often used in poor areas without food

d　国際的な協力に、政府としてではなく、関係する組織のこと

An organization that is not a government associated with international cooperation

8　是正　ぜせい correction

a　問題を解決するための制度を作ること

Creating a system to solve problems

b　仕事のやり方が間違っている時にそれを改めて正しくすること

Renewing and correcting ways of work when they are wrong

c　制度に問題がある時にそれを改めて正しくすること

Renewing and correcting a system when it has a problem

d　施設や場所をよくするために働くこと

Working to make places and facilities better

Talk 2

9　哺乳類　ほにゅうるい mammal

a　ウィルスの病気の患者が、以前、使った物のこと

Things that patients sick with viruses used to use before

b　動物に病気を移す生き物のグループのこと

Groups of living creatures that transmit diseases to animals

c　人などの母乳で育つ動物のグループのこと

Groups of animals that breast-feed, such as humans

d　子どもがよく食べる物のこと

Things that children often eat

10　成虫　せいちゅう imago

a　蚊が病気を移す動物がいっしょにいる虫のこと

Insects that are with animals with diseases transmitted by mosquitoes

b　蚊が水中から出て来る状態で、もう形も変わらないもののこと

the state by which mosquitoes come out of water and their shapes remain unchanged

c　蚊の雄が蜜を取る花の状態のこと

the state of flowers from which male mosquitoes take nectar

d　動物の子どもが自由生活できる状態のこと

the state where young animals can live freely

11　ボウフラ wiggler

a　蚊が大きくなる前の幼虫で、魚のエサのこと

Larvae that are mosquitoes before they grow up and that are fish bait

b　日本にいる花粉を運ぶ虫のこと

Insects in Japan that carry pollen

c　暑い国の病気のこと

Diseases in hot countries

d　午後に花が咲いて満開になる植物のこと

Plant that bloom in full in the afternoon

12　棲息　せいそく inhabitation, living

a　蚊が血を吸うのを決める基準のこと

Criteria that determine blood sucking behaviors of mosquitoes

b　特定の国や地域の病気の存在のこと

Existence of diseases in specified countries and areas

c　生き物が住んでいきていること

Living creatures living

d　生き物の生態を調べること

Investigation of creatures' ecology

13　樹液　じゅえき tree sap

a　魚の体から出て来る水の液体のこと

Watery liquid that seeps from fish bodies

b　血を吸う蚊の雌と雄が入っているところのこと

A place where both female and male blood sucking mosquitoes are located

c　木の幹から出て来る水のような液体のこと

Watery liquid that seeps from a tree trunk

d　木に関係のある病気のこと

A disease associated with trees

14　里山　さとやま woodlands near town

a　雌の蚊が病気を移す時に必ず通るプロセスのこと

Process that female mosquitoes always go through when they transmit diseases

b　シラミやノミのような人に関係している虫のこと

Insects that associate with people, such as lice and fleas

c　日本でニワトリの卵が作られて集められるところ

A place where chicken eggs is produced and gathered in Japan

d　町に近く、人の生活に関係がある山のこと

Woodlands that are associates with peoples' lives close to town

15　吸蜜　きゅうみつ nectar sucking

a　咲いている花の蜜を吸う活動のこと　26

An activity of sucking nectar from blooming flowers

b　人間にとって大変重要な活動の一つ

One of the very important activities for humans

c　ウィルスの代表的なもの

A typical virus

d　動物の子どもの生きている環境を調べるときに使われるもの

Thing used when investigating the living environments of young animals

16　野草　やそう wild grass

a　魚の研究の時に病気がわかるもの　22

thing that reveals a disease when studying fish

b　山や野原などに自然に生える草

naturally grown grass in places, such as mountains and fields

c　蚊が血を吸う時に基準として使っているもの

Criteria used when mosquitoes suck blood

d　他の種類の子どもを食べる草

Grass that eat the youth of other kinds

Talk 3

17　望遠鏡　ぼうえんきょう telescope

a　身体の全体などを写す時に使える鏡のこと

A mirror that you can use when reflecting an entire body

b　空の雲を観測する時に雲を写す装置のこと

A device to reflect clouds when observing clouds in the sky

c　遠くにあるものをレンズで大きくして見る装置のこと

A device that enlarge something in distance with lenses

d　自然を観測するのによい場所のこと

A good place to observe the nature

18　可視光　かしこう visible light

a　空港の荷物検査で使われる光

Light used for the baggage inspection at the airport

b　人の目で普通に見える光

Light that can be seen normally with human eyes

c　ブラックホールが観測できる光
Light that can observe a blackhole
d　普通雲の中からは見えない光
Light that cannot normally be seen in the clouds

19　銀河系　ぎんがけい the Galaxy
a　地球を含む多くの星の集団で、夜空に主に見える星
Group of many stars including the Earth and stars mainly seen in the night sky
b　宇宙線シャワーと雷に関係があると考えている人たちのこと
People who think that there is a relationship between cosmic radiation and lightening
c　携帯電話で使われるエネルギーの高い光のグループのこと
High energy light group used for cell phones
d　静電気の空間からしか観測できない星の集団
Group of stars that can be observed only from space with static electricity

20　太古　たいこ ancient times
a　ガンマ線と関係があった時代のこと
Era that has a relationship with gamma rays
b　雲の中の強い静電気の空間が関係する状態
A state relating to space with strong static electricity in the clouds
c　何百年、何千年も前の大昔
Long time ago, such as hundreds and thousands of years ago
d　地球を取り巻いている大気の層の一部
A part of layer of air surrounding the Earth

21　エニグマ enigma
a　どうして起こるのかなど原因が明らかにされていない謎の問題
A mysterious problem that is not solved a cause, such as why it happens
b　地球の外でだけ観測できるエネルギー
Energy that can be observed only outside the Earth
c　高エネルギーの宇宙線シャワーを速めるもの
Thing that accelerates high energy cosmic radiation
d　静電気の空間がある時に確認できるエネルギー
Energy that can be found when there is space with static electricity

22　放電　ほうでん electric discharge
a　赤外線が原因で起こる現象で、暖かくなる状態
A phenomenon occurring due to infrared radiation and a state that becomes warmer

b　空気を切り裂いて電流が流れるくらい強くなっている静電気の現象
A phenomenon of static electricity that becomes strong to the extent where electricity streams by breaking the air

c　携帯電話の電波と関係がある空気中の電気
Electricity in the air that has a relationship with electric waves for cell phones

d　コンピュータの長時間使用で起こる静電気の現象
A phenomenon of static electricity that occurs due to long time use of a computer

23　バルーン balloon

a　エネルギーの高い低いに関係なく、簡単にエネルギーが観測できるもの
A thing that you can easily observe energy regardless of high and low energy

b　特殊な星を調べられる特別な装置
A special device that can investigate a special star

c　ガンマ線の一種で、静電気の計測ができるもの
A type of gamma rays that can measure static electricity

d　ガスや熱い空気を入れた大きな風船のようなもので、人も乗ることができるもの
A thing like a big balloon with gas and hot air, and where people can ride

24　粒子　りゅうし particle

a　静電気が起こるときに体にあたって痛い小さい物
Small things that hit and cause pain in the body when occurring static electricity

b　エネルギーの光に変わる小さい粒
Small particles that change into light of energy

c　宇宙にある大変小さな粒の物
Very small particles that exist in the outer space

d　水などの液体の世界以外にある小さい粒
Small particles/grains that exist except the world of liquid, such as water

Talk 4

25　黒板　こくばん blackboard

a　料理を作る時に使う黒い板
a black board that you use when you cook.

b　教室でノートの代わりに使われる小さい黒い板
a small black board used in the classroom in replacement of a notebook

c　教室で先生が生徒に教える時に使う黒い板
a black board that teachers use when teaching students in the classroom

d　催しでステージを作る時に便利な黒い板
a convenient black board when making a stage at an event

26　学び舎　　まなびや school
a　塾や予備校のこと
schools, such as prep schools and cram schools
b　子どもが教育を受ける学校の良い言い方
a good way to say a school for children's education
c　子どもの教育用のテストなどを作っている会社のこと
a company that creates tests for children's education
d　教育のために必要な文房具を売っているところ
a place where stationary necessary for education is sold

27　下校　げこう going home after classes
a　授業が全部終わって、学校から家に帰ること
going home from school after all classes are done
b　学校のレベルを下に下げること
dropping a school level lower
c　学校に来て教室ですぐに行う、子どもたちの活動
Children's activities as soon as they come to the classroom in school
d　家で宿題としてできる子どもたちの勉強のこと
Children's study where they can do homework at home

28　ロゴ logo
a　テレビコマーシャルの中で使われる大切な言葉
an important word used in TV commercials
b　学校の中の催しが行える広い場所
a large place in school where events can be held
c　会社や特別なグループのしるしで、デザインなども考えられたマーク
a mark thought up as a design as well as a label for a company or a special group
d　子どものおもちゃで、たてものなどを組み立てるのに使える小さいプラスチックのブロック
small plastic toy blocks for children that can be used to assemble things, such as buildings

29　抜粋　ばっすい excerpt
a　本などから勉強やレポートのために必要なところだけをコピーしたもの
a copied part from a book that is only necessary for study and report

b　本などの報告書のこと
report on something, such as books
c　勉強の内容理解に役に立つ情報の入ったビデオのこと
a video containing useful information for understanding content of study
d　検定教科書を使うのに、保護者のために準備された集まり
a gathering prepared for guardians in order to use approved textbooks

30　ドリル　drill notebooks
a　子どもが勉強する時に聞く教材
listening materials for children's study
b　子どもの勉強のために、特に漢字や算数などの練習をするためのノート
children’s' study notebooks especially for practicing kanji and math
c　教科書ではなくて、一般の人向けの聞く教材
listening materials for general public that are not textbooks
d　子どもたちが自主的な活動のできる特別な部屋のこと
a special room for children's voluntary activities

31　縦割り　たてわり classes with mixed age
a　日本の教室内の机の並べ方で、横よりも縦を大切にする並べ方
Japanese way of organizing desks in the classroom that values a vertical line more than a horizontal line
b　日本の文化で、割りばしを縦に割る割り方
a Japanese cultural way to split disposal wooden chopsticks vertically
c　子どもが並ぶ時に、背の高い子供が前にすわるやり方
a way that tall children sit in front when children line up
d　年齢に関係なく、上級生と下級生が一緒に勉強するクラス
a class where both higher- and lower grade students study together regardless of their age

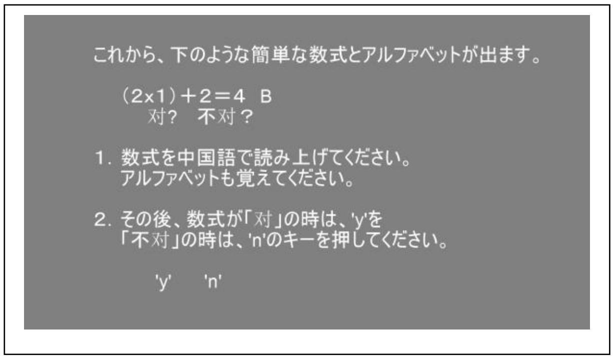
32　推奨　すいしょう recommendation endorsement
a　二つの違うものをつなぐ存在ということ
an existence that connects two different things
b　学校の成績や性格を考えて、コンテストに向いているということ
being well-suited for contests considering school grades and personality
c　よい点を説明して、何かを人に勧めること
recommending something to people after explaining good points
d　人のはっきりとした希望やニーズのこと
people's clear wish and needs

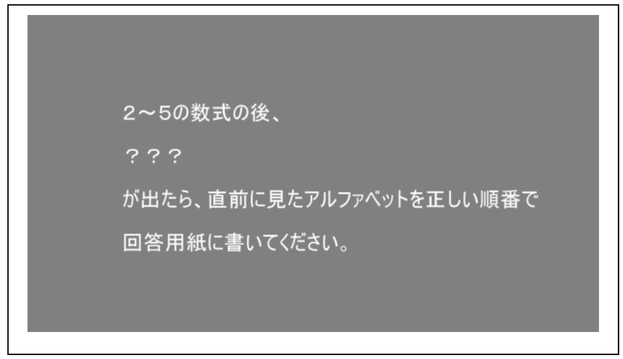
Appendix H. Operation span (OSPAN) task

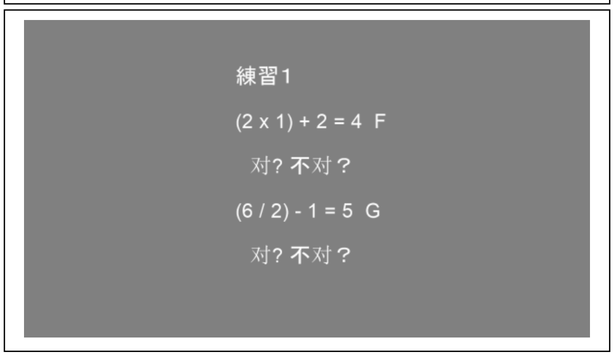
List of Items

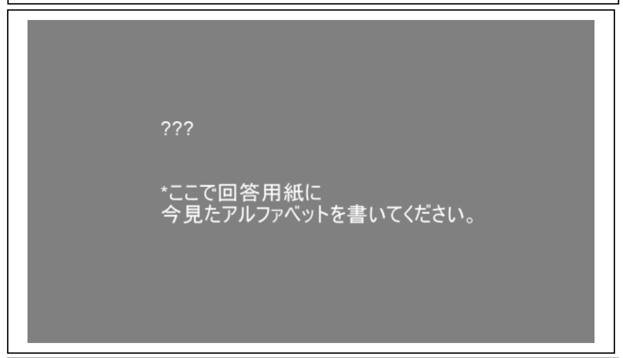
Set *N* = 106	Item		Equation	Correct/ Incorrect	???	Answer
1	1	1311	(3 x 3) – 1 = 8 O	y		
	2	1312	(6/3) + 3 = 2 C	n		
	3	1313	(6/2) – 1 = 2 P	y	OCP	A131
2	1	2511	(4 x 1) – 3 = 1 Z	y		
	2	2512	(6/3) + 4 = 9 P	n		
	3	2513	(4 x 2) – 6 = 2 E	y		
	4	2514	(5/5) + 5 = 6 N	y		
	5	2515	(2 x 4) – 7 = 5 K	n	ZPENK	A251
3	1	3321	(8/4) + 1 = 6 E	n		
	2	3322	(6 x 1) – 2 = 4 T	y		
	3	3323	(4/2) + 3 = 9 Y	n	ETY	A332
4	1	4211	(9/1) – 2 = 5 L	n		
	2	4212	(2 x 3) + 1 = 7 A	y	LA	A421
5	1	5521	(9/3) + 4 = 5 V	n		
	2	5522	(3 x 2) – 5 = 8 D	n		
	3	5523	(6/1) + 2 = 8 O	y		
	4	5524	(2 x 4) – 5 = 3 W	y		
	5	5525	(4/1) + 5 = 3 B	n	VDOWB	A552
6	1	6411	(3 x 3) – 4 = 2 F	n		
	2	6412	(8/2) + 2 = 9 K	n		
	3	6413	(2 x 2) – 3 = 1 L	y		
	4	6414	(2/2) + 3 = 4 R	y	FKLR	A641
7	1	7221	(5 x 2) – 6 = 2 N	n		
	2	7222	(9/3) – 2 = 1 S	y	NS	A722
8	1	8421	(3 x 2) – 4 = 2 L	n		
	2	8422	(8/4) + 6 = 1 F	n		
	3	8423	(2 x 5) – 4 = 6 U	y		
	4	8424	(8/4) + 2 = 9 J	n	LFUJ	A842
9	1	9331	(3 x 2) + 2 = 4 R	n		
	2	9332	(8/2) – 1 = 3 M	y		
	3	9333	(5/5) + 7 = 8 C	y	RMC	A933
10	1	10531	(3 x 3) – 6 = 6 G	n		
	2	10532	(8/4) + 3 = 5 N	y		
	3	10533	(2 x 4) – 7 = 5 X	n		
	4	10534	(4/2) – 1 = 4 F	n		
	5	10535	(2 x 1) + 3 = 5 Y	y	GNXFY	A1053
11	1	11431	(8/2) + 3 = 7 B	y		
	2	11432	(2 x 4) + 1 = 5 N	n		
	3	11433	(7 x 1) – 5 = 2 S	y		
	4	11434	(4 x 2) + 1 = 3 T	n	BNST	A1143

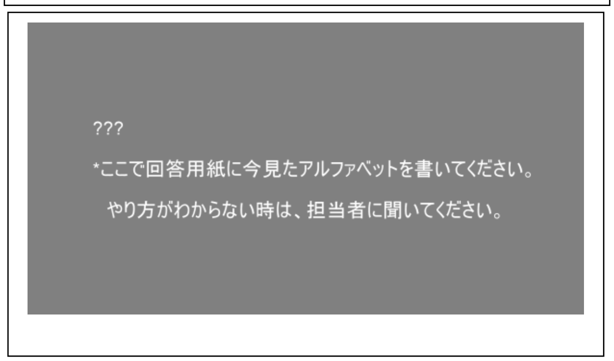
Appendix H. (cont'd) Sample displays of OSPAN

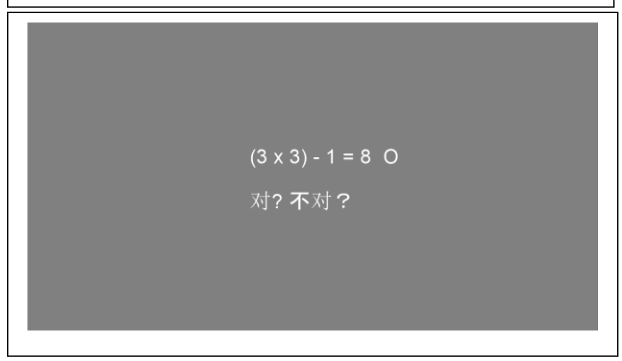






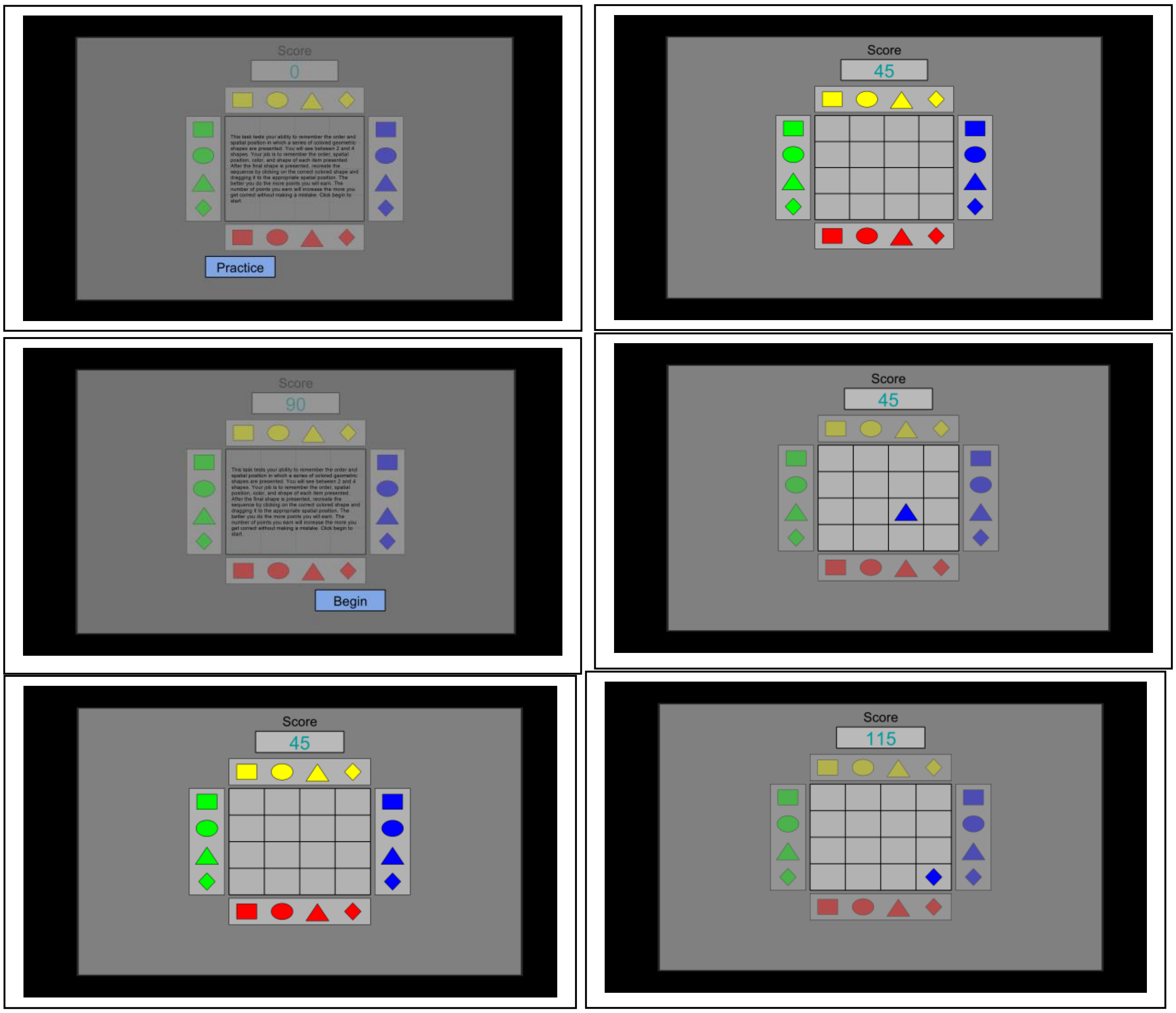






Appendix I. Shapebuilder

Sample displays of Shapebuilder



Appendix J. Length of audio files

			Sec.1	CQ Sec.1	Sec.2	CQ Sec.2	Sec.3	CQ Sec.3	Total seconds	Total minutes
G	Talk 1	AKID	85	120	140	120	114	300	879	14.65
	Talk 2	TOPH	101	120	121	120	155	300	917	15.28
	Talk 3	TEAP	141	150	107	150	141	375	1064	17.73
	Talk 4	MEHL	130	150	161	120	71	300	932	15.53

S	Talk 1	AKID	110	120	188	120	103	300	941	15.68
Talk 2	TOPH	93	120	117	120	140	300	890	14.83	
Talk 3	TEAP	132	150	91	150	131	375	1029	17.15	
Talk 4	MEHL	104	150	152	120	71	300	897	14.95	
E	Talk 1	AKID	121	120	192	120	126	300	979	16.32
Talk 2	TOPH	147	120	169	120	185	300	1041	17.35	
Talk 3	TEAP	155	150	132	150	152	375	1114	18.57	
Talk 4	MEHL	151	150	182	120	89	300	992	16.53	
ME	Talk 1	AKID	134	120	228	120	132	300	1034	17.23
Talk 2	TOPH	157	120	171	120	191	300	1059	17.65	
Talk 3	TEAP	156	150	140	150	159	375	1130	18.83	
Talk 4	MEHL	163	150	190	120	96	300	1019	16.98	
CQ alone										
Talk 1	AKID		120		120		300	240	4.00	
Talk 2	TOPH		120		120		300	240	4.00	
Talk 3	TEAP		150		150		375	300	5.00	
Talk 4	MEHL		150		120		300	270	4.50	
VT	FRT							15x32	480	8.00
MST							40x32	1280	21.00	
MDT							20x32	640	10.00	

Appendix K. Debriefing questionnaire

Debriefing sheet

ID # ______________________________ Pseudonym ______________________________

Thank you for participating in this study. The information you provide here will be kept confidential and will be used only for the purpose of the study.
この研究に参加してくださってありがとうございました。ここでいただいた情報は機密として保管されて、この研究の目的のためだけに使用されます。

*Please answer the following questions. (The researcher will ask a participant these questions in person.)
次の質問に答えてください。（研究者は、これらの質問を直接参加者に尋ねます。）

1. When listening to the talks, which of the following did you focus on?
話を聞いていた時、次のどちらを集中して聞いていましたか。

a. content b. vocabulary
内容 語彙

2. Did you notice words that appeared on the vocabulary tests when you listened to the talks?
話を聞いていた時、語彙テストに出ていたことばに気が付きましたか。

Yes No
はい いいえ

If yes, why did you notice them? (Please be specific.)
はいの場合、どうして気が付きましたか。（詳しく教えてください。）

Appendix L. Results of ANCOVA

Corresponding to 5.5, Group comparisons: Comprehension question scores

Corresponding to 5.5.1, Logistic regressions for main effects of treatment groups

Table L1

Results of Analysis of Covariance for Input Type on CQs

Coefficients	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	8.57	0.18	46.40	0.000***
Simplified	0.29	0.26	1.09	0.278
Elaborated	0.22	0.27	0.80	0.427
Modified Elaborated	0.18	0.26	0.68	0.499
Proficiency	0.67	0.10	6.92	0.000***
WM composite	0.32	0.10	3.22	0.001**

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type/Group	3	9.85	3.28	1.19	0.31
Proficiency	1	177.00	177.00	64.36	0.000***
WM composite	1	28.48	28.48	10.36	0.001**
Residuals	312	858.02	2.75		

ANOVA Type III Test	Sum Sq	DF	*F* value	*p* value
Intercept	5920.2	1	2152.76	0.000***
Input type	3.5	3	0.43	0.73
Proficiency	131.6	1	47.86	0.000***
WM composite	28.5	1	10.36	0.001**
Residuals	858.0	312		

Note. Significant codes are: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘†’ 0.1

Corresponding to 5.5.2, Logistic regressions for main effects of question items

Table L2

Results of Analysis of Covariance for Item Type on CQs

Coefficients	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	8.76	0.16	55.09	0.000***
Synthesis	-0.36	0.22	-1.59	0.112
Inference	0.28	0.22	1.26	0.209
Proficiency	0.67	0.09	7.11	0.000***
WM composite	0.32	0.09	3.40	0.000***

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Item type	2	21.91	10.96	4.08	0.018*
Proficiency	1	180.80	180.80	67.40	0.000***
WM composite	1	30.98	30.98	11.55	0.000***
Residuals	313	839.64	2.68		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	8141.4	1	3034.93	0.000***
Item type	21.9	2	4.08	0.018*
Proficiency	135.5	1	50.49	0.000***
WM composite	31.0	1	11.55	0.000***
Residuals	839.6	313		

Note. Significant codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '†'0.1

Corresponding to 5.5.3, Logistic regressions for main effects of question items by groups

Replication Items

Table L3

Results of Analysis of Covariance for Input Type on CQ Replication Items

Coefficients	*b*	*SE*	*z* value	*p* value
Intercept (Genuine)	8.31	0.31	26.82	0.000***
Simplified	0.60	0.44	1.37	0.175
Elaborated	0.42	0.45	0.93	0.353
Modified Elaborated	0.79	0.44	1.82	0.072†
Proficiency	0.79	0.16	4.85	0.000***
WM composite	0.23	0.17	1.37	0.173

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type	3	13.59	4.53	1.76	0.16
Proficiency	1	74.81	74.81	29.01	0.000***
WM composite	1	4.86	4.86	1.88	0.173
Residuals	312	858.02	2.75		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	1854.13	1	719.07	0.000***
Input type/Group	9.24	3	1.19	0.32
Proficiency	60.61	1	23.51	0.000***
WM composite	4.86	1	1.88	0.17
Residuals	257.85	100		

Note. Significant codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '†' 0.1

Synthesis items in CQ

Table L4

Results of Analysis of Covariance for Input Type on CQ Synthesis Items

Coefficients	*b*	*SE*	*t* value	*p* value
Intercept (Genuine)	8.32	0.31	27.26	0.000***
Simplified	0.26	0.44	0.59	0.555
Elaborated	0.27	0.47	0.58	0.563
Modified Elaborated	0.04	0.43	0.09	0.927
Proficiency	0.52	0.16	3.18	0.002**
WM composite	-0.05	0.25	-0.19	0.849
Simplified x WM composite	1.06	0.42	2.49	0.014*
Elaborated x WM composite	0.14	0.45	0.30	0.762
Modified x WM composite Elaborated	0.78	0.43	1.80	0.074†

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type/Group	3	2.44	0.81	0.33	0.81
Proficiency	1	48.25	48.25	19.34	0.000***
WM composite	1	11.54	11.54	4.62	0.034*
Group x WM composite	3	19.35	6.45	2.59	0.058†
Residuals	97	241.98	2.50		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	1854.06	1	743.22	0.000***
Input type/Group	1.50	3	0.20	0.896
Proficiency	25.17	1	10.09	0.002**
WM composite	0.09	1	0.04	0.849
Group x WM composite	19.35	3	2.59	0.058^
Residuals	241.98	97		

Note. Significant codes are: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘†’ 0.1

Inference items in CQ

Table L5

Results of Analysis of Covariance for Input Type on CQ Inference Items

Coefficients	*b*	*SE*	*t* value	*p* value
Intercept (Genuine)	9.06	0.35	25.82	0.000***
Simplified	-0.05	0.52	-0.10	0.920
Elaborated	0.03	0.53	0.06	0.955
Modified Elaborated	0.11	0.52	0.22	0.829
Proficiency	0.94	0.33	2.82	0.006**
WM composite	0.34	0.30	1.14	0.258
Simplified x Proficiency	-0.33	0.48	-0.69	0.494
Elaborated x Proficiency	-0.20	0.56	-0.36	0.717
Modified Elaborated x Proficiency	-0.67	0.50	-1.35	0.181
Simplified x WM composite	-0.05	0.51	-0.09	0.925
Elaborated x WM composite	0.16	0.52	0.32	0.751
Modified Elaborated x WM composite	0.32	0.53	0.61	0.543
Proficiency x WM composite	0.05	0.25	0.21	0.832
Simp x Prof x WM	0.31	0.43	0.71	0.479
Elab x Prof x WM	-0.27	0.67	-0.40	0.689
Modified x Prof x WM	-0.73	0.48	-1.51	0.135

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type	3	3.39	1.13	0.36	0.79
Proficiency	1	55.47	55.47	17.43	0.000***
WM composite	1	13.27	13.27	4.17	0.044*
Group x Prof	3	4.99	1.66	0.52	0.668
Group x WM	3	0.95	0.32	0.10	0.960
Prof x WM	1	0.05	0.05	0.02	0.901
Gr x Prof x WM	3	12.32	4.11	1.29	0.282
Residuals	90	286.33	3.18		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	2120.92	1	666.65	0.000***
Input type/Group	0.31	3	0.03	0.992
Proficiency	25.30	1	7.95	0.006**
WM composite	4.12	1	1.30	0.258
Group x Prof	5.94	3	0.62	0.602
Group x WM	1.65	3	0.17	0.915
Prof x WM	0.14	1	0.05	0.832
Gr x Prof x WM	12.32	3	1.29	0.282
Residuals	286.33	90		

Note. Significant codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '†' 0.1

Corresponding to 5.6, Group comparisons: IVL
Corresponding to 5.6.1, One factor analysis of covariance (ANCOVA) for main effects of treatment groups on FRT

Table L6

Results of Analysis of Covariance for Input Type on FRT

Coefficients	*b*	*SE*	*t* value	*p* value
Intercept (Genuine)	21.04	0.80	26.43	0.000***
Simplified	-0.36	1.14	-0.31	0.755
Elaborated	2.02	1.15	1.75	0.083†
Modified Elaborated	0.99	1.12	0.88	0.380
Proficiency	2.66	0.42	6.36	0.000***
WM composite	0.26	0.63	0.41	0.684
Simplified x WM	0.57	1.08	0.53	0.595
Elaborated x WM	0.66	1.13	0.58	0.561
Modified x WM Elaborated	1.86	1.13	1.65	0.103

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type	3	122.00	40.67	2.40	0.073†
Proficiency	1	895.13	895.13	52.65	0.000***
WM	1	67.33	67.33	3.96	0.049*
Group x WM	3	46.22	15.41	0.91	0.441
Residuals	97	1649.09	17.00		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	11878.4	1	698.69	0.000***
Input type/Group	86.5	3	1.70	0.173
Proficiency	686.9	1	40.40	0.000***
WM	2.8	1	0.17	0.684
Group x WM	46.2	3	0.91	0.441
Residuals	1649.1	97		

Note. Significant codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '†' 0.1

Corresponding to 5.6.2, One factor analysis of covariance (ANCOVA) for main effects of treatment groups on MST

Table L7

Results of Analysis of Covariance for Item Type on MST

Coefficients	*b*	*SE*	*t* value	*p* value
Intercept (Genuine)	15.50	0.99	15.64	0.000***
Simplified	0.45	1.42	0.32	0.753
Elaborated	2.45	1.41	1.74	0.085†
Modified Elaborated	0.70	1.45	0.48	0.630
Proficiency	2.50	0.95	2.64	0.010**
WM	0.79	0.78	1.01	0.315
Simplified x Proficiency	1.56	1.33	1.17	0.246
Elaborated x Proficiency	1.81	1.55	1.17	0.246
Modified Elaborated x Proficiency	0.52	1.42	0.37	0.714
Simplified x WM	0.26	1.35	0.20	0.846
Elaborated x WM	-0.57	1.38	-0.41	0.680
Modified Elaborated x WM	2.90	1.76	1.65	0.102
Proficiency x WM	0.50	0.67	0.76	0.451
Simp x Prof x WM	1.61	1.20	1.34	0.185
Elab x Prof x WM	-1.51	2.34	-0.65	0.519
Modified x Prof x WM	-0.76	1.34	-0.57	0.572

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type/Group	3	97.27	32.42	1.33	0.269
Proficiency	1	1503.04	1503.04	61.71	0.000***
WM	1	118.72	118.72	4.87	0.030*
Group x Proficiency	3	43.68	14.56	0.60	0.618
Group x WM	3	118.97	39.66	1.63	0.188
Proficiency x WM	1	47.55	47.55	1.95	0.166
Grp x Prof x WM	3	81.18	27.06	1.11	0.349
Residuals	90	2192.00	24.36		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	5958.5	1	244.65	0.000***
Group	84.3	3	1.15	0.332
Proficiency	169.1	1	6.94	0.010**
WM	24.8	1	1.02	0.315
Group x Proficiency	50.4	3	0.69	0.560
Group x WM	84.4	3	1.16	0.331
Proficiency x WM	14.0	1	0.57	0.451
Grp x Prof x WM	81.2	3	1.11	0.349
Residuals	2192.4	90		

Note. Significant codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '† '0.1

Corresponding to 5.6.3, One factor analysis of covariance (ANCOVA) for main effects of treatment groups on MDT

Table L8

Results of Analysis of Covariance for Input Type on MDT

Coefficients	*b*	*SE*	*t* value	*p* value
Intercept (Genuine)	18.24	0.81	22.49	0.000***
Simplified	1.78	1.15	1.54	0.126
Elaborated	3.18	1.16	2.73	0.007**
Modified Elaborated	0.95	1.14	0.84	0.405
Proficiency	2.17	0.77	2.81	0.006**
WM composite	1.13	0.42	2.68	0.009**
Simplified x Proficiency	1.96	1.07	1.83	0.070†
Elaborated x Proficiency	1.49	1.28	1.17	0.245
Modified Elaborated x Proficiency	0.90	1.10	0.82	0.414

ANCOVA	*DF*	Sum Sq	Mean Sq	*F* value	*p* value
Input type/Group	3	175.47	58.49	3.35	0.022*
Proficiency	1	1281.00	1281.00	73.41	0.000***
WM	1	118.49	118.49	6.79	0.011*
Group x Proficiency	3	62.45	20.82	1.19	0.317
Residuals	97	1692.53	17.45		

ANOVA Type III Test	Sum Sq	*DF*	*F* value	*p* value
Intercept	8826.9	1	505.88	0.000***
Group	140.5	3	2.68	0.051†
Proficiency	137.3	1	7.87	0.006**
WM	125.5	1	7.19	0.009**
Group x Proficiency	62.4	3	1.19	0.317
Residuals	1692.5	97		

Note. Significant codes are: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘† ’0.1

Appendix M. Correlation table [Pearson] (*N* = 106)

	Grp	CQ	FRT	MST	MDT	Prof	OS	SB	Focus	Notice	Age	Job	LOR	FIJ	AFI
Group	1	.080	.150	.067	.117	.094	.052	-.019	.014	-.249*	-.076	-.016	-.033	.063	-.086
Comp. Q	.080	1	.627**	.761**	.700**	.539**	.240*	.294**	-.132	-.220*	.022	-.283**	.180	.084	-.284**
Form Recog. T	.150	.627**	1	.756**	.790**	.600**	.090	.297**	-.025	-.265**	.221*	-.341**	.344**	.050	-.238*
M-R sentence T	.067	.761**	.756**	1	.841**	.607**	.153	.313**	-.105	-.172	.107	-.246*	.253**	.022	-.317**
M-R def. T	.117	.700**	.790**	.841**	1	.638**	.155	.339**	-.090	-.202*	.078	-.249**	.221	-.063	-.185
Proficiency T	.094	.539**	.600**	.607**	.638**	1	.127	.226*	.011	-.131	-.116	-.109	.069	.218*	-.327**
OSPAN (WM)	.052	.240*	.090	.153	.155	.127	1	.202*	-.091	.061	-.095	.049	-.160	.003	.091
Shapebuil (WM)	-.019	.294**	.297**	.313**	.339**	.226*	.202*	1	-.085	-.032	.037	-.081	.109	-.088	.064
Focused on content	.014	-.132	-.025	-.105	-.090	.011	-.091	-.085	1	-.006	-.018	.003	-.014	.125	-.054
Noticed TW	-.249*	-.220*	-.265**	-.172	-.202*	-.131	.061	-.032	-.006	1	-.084	.011	-.073	-.006	.195*
Age	-.076	.022	.221*	.107	.078	-.116	-.095	.037	-.018	-.084	1	-.608**	.622**	.002	.256**
Job	-.016	-.283**	-.341**	-.246*	-.249**	-.109	.049	-.081	.003	.011	-.608**	1	-.573**	-.124	.026
Length of resi.	-.033	.180	.344**	.253**	.221*	.069	-.160	.109	-.014	-.073	.622**	-.573**	1	-.133	-.014
Formal instruct.	.063	.084	.050	.022	-.063	.218*	.003	-.088	.125	-.006	.002	-.124	-.133	1	-.438**
Age of onset of FI	-.086	-.284**	-.238*	-.317**	-.185	-.327**	.091	.064	-.054	.195*	.256**	.026	-.014	-.438**	1

* Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

Note. "Job" includes 1. Professional; 2. Graduate student; 3. Undergraduate student; 4. Language school student; 5. Other.

Appendix N. Logistic MEM model-building process

Interaction examination with CQ

1 CQ binary response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
2 CQ binary response ~ Input Type* composite WM + (1|Prsn) + (1|Item)
3 CQ binary response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item)
4 CQ binary response ~ Item Type*centered Prof + (1|Prsn) + (1|Item)
5 CQ binary response ~ Item Type* composite WM + (1|Prsn) + (1|Item)
6 CQ binary response ~ Item Type* centered Prof * composite WM + (1|Prsn) + (1|Item)

7 CQ binary Replication response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
8 CQ binary Replication response ~ Input Type* composite WM + (1|Prsn) + (1|Item)
9 CQ binary Replication response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item)

10 CQ binary Synthesis response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
11 CQ binary Synthesis response ~ Input Type* composite WM + (1|Prsn) + (1|Item) <Interaction was found>
12 CQ binary Synthesis response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item)

13 CQ binary Inference response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
14 CQ binary Inference response ~ Input Type* composite WM + (1|Prsn) + (1|Item)
15 CQ binary Inference response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item) <Interaction was found>

CQ: Input Type

1 CQ binary response ~ Input Type + (1|Prsn) + (1|Item)
2 CQ binary response ~ Input Type + centered Prof + (1|Prsn) + (1|Item)
3 CQ binary response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item) <Best Fitting Model>

CQ: Item Type

1 CQ binary response ~ Item Type + (1|Prsn) + (1|Item)
2 CQ binary response ~ Item Type + centered Prof + (1|Prsn) + (1|Item)
3 CQ binary response ~ ItemType + centered Prof + composite WM + (1|Prsn) + (1|Item) <Best Fitting Model>

CQ: Replication

1 CQ binary Replication response ~ Input Type + (1|Prsn) + (1|Item)
2 CQ binary Replication response ~ Input Type + centered Prof + (1|Prsn) + (1|Item)
3 CQ binary Replication response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item) <Best Fitting Model>

CQ: Synthesis

1 CQ binary synthesis response ~ Input Type + (1|Prsn) + (1|Item)
2 CQ binary Synthesis response ~ Input Type + centered Prof + (1|Prsn) + (1|Item)
3 CQ binary Synthesis response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item)
4 CQ binary Synthesis response ~ Input Type + centered Prof + composite WM + Input type*composite WM + (1|Prsn) + (1|Item) <Best Fitting Model>

CQ: Inference

1 CQ binary Inference response ~ Input Type + (1|Prsn) + (1|Item)
2 CQ binary Inference response ~ Input Type + centered Prof + (1|Prsn) + (1|Item)
3 CQ binary Inference response ~ Input Type + centered Prof + composite WM + (1|Prsn) + (1|Item)
4 CQ binary Inference response ~ Input Type + centered Prof + composite WM + Input Type*centered Prof*composite WM + (1|Prsn) + (1|Item) <Best Fitting Model>

Interaction examination with VT

1 FRT binary response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
2 FRT binary response ~ Input Type* composite WM + (1|Prsn) + (1|Item) <Interaction was found>
3 FRT binary response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item)

7 MST binary response ~ Input Type* centered Prof + (1|Prsn) + (1|Item)
8 MST binary response ~ Input Type* composite WM + (1|Prsn) + (1|Item) <Interaction was found>
9 MST binary response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item) <Interaction was found>

10 MDT binary response ~ Input Type* centered Prof + (1|Prsn) + (1|Item) <Interaction was found>
11 MDT binary response ~ Input Type* composite WM + (1|Prsn) + (1|Item)

12 MDT binary response ~ Input Type* centered Prof * composite WM + (1|Prsn) + (1|Item)

VT: FRT

1 FRT binary response ~ Input Type + (1|Prsn) + (1|TW)

2 FRT binary response ~ Input Type + centered Prof + (1|Prsn) + (1|TW)

3 FRT binary response ~ Input Type + centered Prof + centered WM(SB) + (1|Prsn) + (1|TW)

4 FRT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered Prof + (1|Prsn) + (1|TW)

5 FRT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB) + (1|Prsn) + (1|TW) <Best Fitting Model>

VT: MST

1 MST binary response ~ Input Type + (1|Prsn) + (1|TW)

2 MST binary response ~ Input Type + centered Prof + (1|Prsn) + (1|TW)

3 MST binary response ~ Input Type + centered Prof + centered WM(SB) + (1|Prsn) + (1|TW)

4 MST binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB) + (1|Prsn) + (1|TW)

5 MST binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB)*centered Prof + (1|Prsn) + (1|TW) <Best Fitting Model>

VT: MDT

1 MDT binary response ~ Input Type + (1|Prsn) + (1|TW)

2 MDT binary response ~ Input Type + centered Prof + (1|Prsn) + (1|TW)

3 MDT binary response ~ Input Type + centered Prof + centered WM(SB) + (1|Prsn) + (1|TW)

4 MDT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered Prof + (1|Prsn) + (1|TW) <Best Fitting Model>

5 MDT binary response ~ Input Type + centered Prof + centered WM(SB) + Input Type*centered WM(SB) + (1|Prsn) + (1|TW)

List of References

Al-Thowaini, A. (2018). *Speech modification to non-native speakers and content dilution: Implications for English as a medium of instruction (EMI)* (Unpublished doctoral dissertation). University of Maryland, College Park.

Atkins, S. M., Sprenger, A. M., Colflesh, G. J. H., Briner, T. L., Buchanan, J. B., Chavis, S. E., Chen, S., Iannuzzi, G. L., Kashtelyan, V., Dowling, E., Harbison, J. I., Bolger, D. J., Bunting, M. F., & Dougherty, M. R. (2014). Measuring working memory is all fun and games: A four-dimensional spatial game predicts cognitive task performance. *Experimental Psychology, 61*(6), 417–438. https://doi.org/10.1027/1618-3169/a000262

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208.

Baddeley, A. (2015). Working memory in second language learning. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Bristol, UK: Multilingual Matters.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Barcroft, J. (2015). Can retrieval opportunities increase vocabulary learning during reading? *Foreign Language Annals, 48*(2), 236–249. https://doi.org/10.1111/flan.12139

Barcroft, J. (2016). *Vocabulary in language teaching*. New York: Routledge.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2016). lme4: Linear mixed-effects models using "Eigen" and S4. R package (version 1.1–12) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/lme4/index.html

Bisson, M-J., Van Heuven, W. J. B., Cocklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *The Quarterly Journal of Experimental Psychology, 68*(7), 1306–1326. https://doi.org/10.1080/17470218.2014.979211

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension: A technical report.* Retrieved from http://www.casl.umd.edu/node/127

Borro, I. (2020). *Enhanced incidental learning of formulaic sequences by Chinese learners of Italian* (Unpublished doctoral dissertation). University of Portsmouth.

Breiner-Sanders, K. E., Swender, E., & American Council on the Teaching of Foreign Languages. (1999). *The ACTFL Oral Proficiency Interview: Tester training manual.* Yonkers, NY: American Council on the Teaching of Foreign Languages.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the revised hierarchical model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13*(3), 359–371.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Chaudron, C. (1983). Simplification of input: Topic reinstatements and their effects on L2 learners' recognition and recall. *TESOL Quarterly, 17*(3), 437–458.

Chaudron, C. (1985a). Comprehension, comprehensibility, and learning in the second language classroom. *Studies in Second Language Acquisition, 7*(2), 216–232.

Chaudron, C. (1985b). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition, 7*(1), 1–14.

Chaudron, C., & Richards, J. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, *7*(2), 113–127.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031

Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*(2), 316–325.

Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly, 26*(2), 345–374.

Chung, H. (1995). Effects of elaborative modifications on second language reading comprehension and incidental vocabulary learning. *University of Hawai'i Working Papers in ESL, 14*(1), 27–61.

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01345

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.

Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords.* Bristol, UK: Multilingual Matters.

Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *Journal of Experimental Education, 56*(2), 67–76.

Derwing, T. M. (1996). Elaborative detail: Help or hindrance to the NNS listener? *Studies in Second Language Acquisition, 18*(3), 283–297.

De Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhofer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning, 68*(4), 906–941. https://doi.org/10.1111/lang.12296

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23*(2), 166–180.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in*

Second Language Acquisition, 40(2), 341–366. https://doi.org/10.1017/S0272263117000109

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174–187.

Ford-Niwa, J., Kobayashi, N., & Yamamoto, H. (1995). *Nihongo nooryoku kani shaken (SPOT) wa nani o sokutei shiteiru ka: onsei teepu yooin no kaiseki* [What does a “Simple Performance-Oriented Test (SPOT)” measure? A study of the tape listening factor]. *Nihongo kyoiku* [Japanese language education], *86*, 93–102.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*(1), 65–77.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H-J. (2017). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

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. https://doi.org/10.1017/S0272263113000119

González-Fernández, B., & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen & M. Sato (Eds.), *Routledge handbook of instructed second language acquisition* (pp. 280–298). New York: Routledge.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, *61*(4), 481–502.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25–38.

Hatami, S. (2017). The differential impact of reading and listening on L2 language acquisition of different dimensions of word knowledge. *Reading in a Foreign Language*, *29*(1), 61–85.

Hatasa, Y. A., & Tohsaku, Y. (1997). SPOT as a placement test. In H. Cook, K. Hijirida, & M. Tahara (Eds.), *New trends and issues in teaching Japanese language and culture* (Technical Report Series, no. 15, pp. 77–98). Honolulu: National Foreign Language Resource Center, University of Hawaiʻi.

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00017

Hulme, R. C., Barsky, D., & Rodd, J. M. (2018). Incidental learning and long-term retention of new word meaning from stories: The effect of number of exposures. *Language Learning*, *69*(1), 18–34. https://doi.org/10.1111/lang.12313

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell.

Iori, I., Iwata, K., & Mori, A. (2011). Rewriting official documents into "easy Japanese." *Jinbun shizenkenkyuu* (Research on Humanities and Nature: Departmental Bulletin, no. 5, pp. 115–139). Hitotsubashi University.

Issa, B. I., & Morgan-Short, K. (2018). 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

The Japan Foundation/Japan Educational Exchanges and Services. (2012). *N1–N5: Summary of linguistic competence for each level*. Retrieved August 1, 2020, from https://www.jlpt.jp/e/about/levelsummary.html

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47–77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24*(4), 617–637.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416–432.

Jiang, N. (2018). *Second language processing: An introduction*. New York: Routledge.

Kim, Y. (2006). Effects of input elaboration on vocabulary acquisition through reading by Korean learners of English as a foreign language. *TESOL Quarterly, 40*(2), 341–373.

Kobayashi, H., & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning, 42*(2), 183–215.

Kobayashi, N., Sakai, T., & Ford-Niwa, J. (2007). *Sokutoo yookyuugata gengo test no WEB-ka: SPOT-WEB no baai* [A WEB-based quick response language test: The SPOT-WEB test]. *CASTEL-J in Hawaii 2007 Proceedings* (pp. 231–234). Retrieved November 7, 2020 from http://castelj.kshinagawa.com/proceedings/files/3-C4%20Kobayashi.pdf

Koda, K. (2005). Information integration in sentence processing. In K. Koda (Ed.), *Insights into second language reading: A cross-linguistic approach* (pp. 95–153). Cambridge: Cambridge University Press.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261–271.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (Vol. 2, pp. 389–395). London: Wiley.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174.

Kroll, J. F., & Sundermant, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representaitons. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 104–129). Oxford: Blackwell.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*(3), 373–381.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition*. London: Longman.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26.

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *Routledge handbook of second language acquisition* (pp. 163–176). London: Routledge.

Lee, J. (2016). *Nihongo kyoiku no tame no bunshoo nanido kenkyu* [Research on text difficulty for Japanese language education], *Waseda nihongo kyoikugaku* [Waseda Japanese language education], *21*, 1–16. Waseda University Graduate School Japanese Language Education Program.

Lee, S., & Huang, H. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition, 30*(3), 307–331. https://doi.org/10.10170S0272263108080479

Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of the L2. In S. M. Gass, P. S. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 167–186). New York: Routledge.

Linck, J., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(1), 185–207. https://doi.org/10.1111/lang.12117

Linck, J., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang12011

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4) 861–883. https://doi.org/10.3758/s13423-013-0565-2

Long, M. H. (1980). *Input, interaction and second language acquisition* (Unpublished doctoral dissertation). University of California, Los Angeles.

Long, M. H. (1981). Input, interaction and second language acquisition. In H. Wintz (Ed.), *Native language and foreign language acquisition: Annals of the New York Academy of Sciences*, *379*, 259–278.

Long, M. H. (1983a). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition*, *5*(2), 177–193.

Long, M. H. (1983b). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*(2), 126–141.

Long, M. H. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Cambridge, MA: Newbury House.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bahtia (Eds.), *Handbook of second language acquisition* (pp. 413– 468). New York: Academic Press.

Long, M. H. (2007) Texts, tasks, and the advanced learner. In M. H. Long (Ed.), *Problems in SLA* (pp. 119–138). New York: Lawrence Erlbaum.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Long, M. H. (2017). ISLA: Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition, 1*(1), 7–44.

Long, M. H. (2019). Optimal input for language teaching: Genuine, simplified, elaborated, or modified elaborated? *Language Teaching, 53*(2), 169–182. https://doi.org/10.1017/S0261444819000466

Long, M. H., & Ross, S. (1993). Modifications that preserve language and content. In M. Tickoo (Ed.), *Simplification: Theory and application* (pp. 29–52). Singapore: SEAMEO Regional Language Centre.

Long, M. H., & Ross, S. (2009). Input elaboration: A viable alternative to “authentic” and simplified texts. In K. Namai & Y. Fukada (Eds.), *Toward the fusion of language, culture and education: From the perspectives of international and interdisciplinary research. A festschrift for Yasukata Yano* (pp. 307–325). Tokyo: Kaitakusha.

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts.* New York: Routledge.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Second Language Acquisition, 16*(3), 303–323.

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition, 40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*(5–6), 383–388.

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3–4), 422–429.

Matsumoto, H. (1989/1994). *Keesu7 Goshu* (*Tango no shakuyoo*) [Case 7 word kinds (loanwords)]. In Y. Morita, S. Muraki, & M. Aizawa (Eds.), *Nihongo no goi* [Japanese vocabulary] (pp. 42–49). Tokyo: Oohuu.

Matsushita, T. (2011). Vocabulary database for reading Japanese (VDRJ), Ver. 1.1. http://www.wa.commufa.jp/~tatsum/

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (rev. ed.). Cambridge: Cambridge University Press.

Nguyen, C. D., & Boers, F. (2019). The effect of content retelling on vocabulary uptake from a TED talk. *TESOL Quarterly*, *53*(1), 5–29.

O'Donnell, M. E. (2009). Finding middle ground in second language reading: Pedagogic modifications that increase comprehensibility and vocabulary acquisition while preserving authentic text features. *The Modern Language Journal, 93*(4), 512–533.

Oh, S.-Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, *35*(1), 69–96.

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*(4), 287–289.

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society, 31*(2), 117–118.

Parker, K., & Chaudron, C. (1987). The effects of linguistic simplification and elaborative modification on L2 comprehension. *University of Hawaiʻi Working Papers in ESL, 6*(2), 107–133.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition, 38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez. (Eds.), *Understanding formulaic language* (pp. 153–173). New York: Routledge.

Peng, C-Y. Y., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Education Research, 96*(1), 1–14.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 23–38.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Reder, L. M. (1982). Elaborations: When do they help and when do they hurt? *Text: Interdisciplinary Journal for the Study of Discourse, 2*(1–3), 211–223.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition, 35*(1), 31–65.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(4), 589–619. https://doi.org/10.1017/S0272263199004039

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78. https://doi.org/10.1016/0346-251X(78)90027-1

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmidt, R. W. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*(2), 165–179.

Shibatani, M. (1990). *The languages of Japan* (rev. ed.). Cambridge: Cambridge University Press.

Suzuki, Y. (2014). A practical method of measuring automaticity in second language grammar: A comparison between web-based SPOT and ACTFL OPI. *The Institute of Electronic, Information, and Communication Engineering (IEICE) Technical Report, 114*(100), 49–54.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & G. Seidohofer (Eds.), *Principles and practices in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45–58.

Tamaru, Y., & Yoshioka, K. (1994). *Nihongo hatsuwa shiryoo bunseki no tan'i o meguru mondai: daini gengo shuutoku katee kansatsu no tachiba kara* [Problems surrounding the units of analyses for spoken data of Japanese as a second language]. *Working Papers, Japanese Language Program, International University*, 84–100.

Tanaka, H., & Mino, H. (2010). Manual translation experiment of broadcast news in simple Japanese. *Information Processing Society of Japan (IPSJ) SIG Technical Report*, *2010-NL-199*, 11. Retrieved November 7, 2020 from https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=71035&item_no=1&page_id=13&block_id=8

Tomlin, R. T., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203.

Toya, M. (1992). *Form of explanation in modification of listening input in L2 vocabulary learning* (Unpublished master's thesis). University of Hawai‘i, Honolulu.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599, https://doi.org/10.1111/lang.12343

Urano, K. (2000). *Lexical simplification and elaboration: Sentence comprehension and incidental vocabulary acquisition* (Unpublished master's thesis). University of Hawai'i, Honolulu.

Vafaee, P. (2016). *The relative significance of syntactic knowledge and vocabulary knowledge in second language listening comprehension* (Unpublished doctoral dissertation). University of Maryland, College Park.

Van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual richness and word learning: Context enhances comprehension but retrieval enhances retention. *Language Learning*, *68*(2), 546–585. https://doi.org/10.1111/lang.12285

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning, 65*(2), 390–416.

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition, *Language Learning, 61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130–163.

Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*(3), 287–307.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52.

Webb, S. (2007a). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2007b). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language, 19*(2), 120–136.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

Wen, Z. (E.) & Li, S. (2019). Working memory in L2 learning and processing. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning.* Cambridge: Cambridge University Press.

Widdowson, H. G. (1976). The authenticity of language data. In J. F. Fanselow & R. Crymes (Eds.), *On TESOL ’76* (pp. 261–270). Washington DC: TESOL.

Winter, B. (2019). *Statistics for linguists: An introduction using R.* London: Routledge.

Wright, C. (2009). The role of working memory in the development of L2 grammatical proficiency. In A. Benati (Ed.), *Issues in second language proficiency* (pp. 45–60). London: Bloomsbury.

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, *44*(2), 189–219.