DO RECASTS PROVIDE SECOND LANGUAGE LEARNERS WITH NEGATIVE EVIDENCE?

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education

By
Hideki Sakai
July, 2015

Examining Committee Members

David Beglar, Advisory Chair, Teaching and Learning
Tomoko Nemoto, Teaching and Learning
Mark Sawyer, External Member, Kwansei Gakuin University
Steven Ross, External Member, University of Maryland
James Elwood, External Member, Meiji University

ABSTRACT

The purpose of this experimental study is to examine the effects of recasts on narrowing overgeneralized grammar in the second language (L2). The study involved testing three major hypotheses of the mechanisms underlying recasts: the direct contrast hypothesis (Saxton, 1997, 2000), the additional input hypothesis (Gass, 1997; Gass & Mackey, 2007; Long, 1996, 2007), and the enhanced salience hypothesis (Leeman, 2003). Two structures (adjective ordering and indirect passives) were selected for this study, mainly because it was assumed that Japanese learners of English might produce overgeneralized rules that allow incorrect structures because of their first language (L1) influence. The participants were 97 Japanese university students learning English as a foreign language in Japan. They were randomly assigned to four treatment groups: recast (the REC Group), non-contingent positive evidence (the POS Group), recast plus additional input (the REC+ Group), and input with enhanced salience (the SAL Group). A pretest, posttest, and delayed-posttest design were employed. The measurement instruments were an oral production task, elicited imitation task, and untimed grammaticality judgment task, each of which was designed to elicit participants' implicit and explicit knowledge about adjective ordering and indirect passives. Thus, the independent variable was the treatment conditions, and the dependent variable were the test scores regarding ungrammaticality of the overgeneralized rules of the target structures. After data screening, the data from 75 of the 97 participants were analyzed for adjective ordering, and the data from 90 participants were analyzed for indirect passives. The results showed that the POS Group did not improve on all the measures for adjective ordering and indirect passives; thus, it was suggested that

positive evidence was not sufficient for the participants to narrow overgeneralized rules for the target structures. The findings indicated that for adjective ordering, medium effect sizes for the comparison of the POS and REC Groups were obtained on the grammaticality judgment tests for the pretest-posttest and pretest-delayed posttest comparisons. Thus, based on these effect sizes, the direct contrast hypothesis was partially supported (i.e., for one of the two structures and one measure of the three tests). Furthermore, on the basis of the results that the REC and REC+ Groups did not differ significantly on any measure and that on the adjective-ordering grammaticality judgment tests, the comparison between the REC+ Group and the POS Group obtained a medium effect size for the pretest-delayed posttest comparison, the provision of recasts in the REC and REC+ Groups was effective at least for the adjective-ordering grammaticality judgment tests; however, additional input alone did not have an impact on L2 learners' retreat from the overgeneralized rules. Finally, the results showed that the POS and SAL Groups did not differ significantly and that there existed a difference in the performance on the adjective-ordering grammaticality judgment tests between the REC Group and the SAL Groups. Salience might not be effective in helping L2 learners retreat from overgeneralized rules, and the enhanced salience hypothesis can be interpreted as being limited to learning some linguistic structures. In conclusion, the present study provides empirical evidence that L2 learners can persist in using incorrect overgeneralized rules due to L1 rules and that ten tokens of the target structures are not effective for L2 learners to retreat from the overgeneralized rules. Furthermore, the findings lend support to the direct contrast hypothesis as a mechanism underlying recasts for one of the two target structures.

ACKNOWLEDGMENTS

I owe a great debt of gratitude to a number of people who were directly or indirectly involved in this research.

First and foremost, I would like to sincerely thank the chair of my dissertation defense committee, Dr. David Beglar, who provided constructive feedback and insightful comments that helped to improve many aspects of this dissertation. I could not have completed this project without his constant encouragement.

I would also like to express my gratitude to Drs. Mark Sawyer, Steven Ross, Tomoko Nemoto, and James Elwood who provided valuable advice and helped me revise the final draft of my dissertation. Particularly, Dr. Mark Sawyer offered much encouragement in the early stages of this research.

Finally I would like to thank Colleen Dalton, Brian Wistner, Nobuhiko Okumura, Junichi Wada, Daisuke Takeuchi, Naoki Tokuda, Kentaro Uchida, Yu Miyazawa, Yu Zhou, Yoko Maruyama, and Tomoyuki Abe for their contribution to this study, specifically in the material development, data collection, and data analysis phases.

Without the kind support from all of the above people, this project would never have reached completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Background of the Issue

Negative feedback to second language (L2) learners' errors has drawn a great deal of attention in the field of L2 teaching and acquisition, and its roles have been differently valued. In the Audiolingual method, which was based on structural linguistics and behavioral psychology, it was argued that the L2 learning process depends on mechanical habit formation and that errors should be avoided so that good habits are formed (Richards & Rodgers, 2001, p. 57; Rivers, 1968, p. 76). Thus, negative or positive feedback should be given immediately after students' performance in order to reinforce the correct behaviors and eliminate erroneous ones (Richards & Rodgers, 2001, p. 64). Later, because L2 acquisition researchers identified developmental errors that all L2 learners produce and that reflect the development of L2 knowledge (e.g., Dulay, Burt, & Krashen, 1982), in L2 teaching approaches and methods, such as the natural approach and communicative language teaching, errors were not viewed as a problem that must be corrected immediately; rather, they were considered to indicate L2 development. The meaning or communicative message took a primary role in those newer approaches and methods, and linguistic form or accuracy was deemphasized. In other words, the methodology shifted from a focus on forms to a focus on meaning (Long, 1991; Long & Robinson, 1998). For example, Littlewood (1981), one of the proponents of communicative language teaching, stated that learners' errors should be overlooked because they reflect the natural progress in the development of communicative competence (p. 94). Similarly, in the natural approach, Krashen and Terrell (1983) argued against the use of overt and direct error corrections because they exert a negative effect on L2 learners' affect and their benefits might be limited to conscious learning. Thus, in these approaches and methods, negative feedback was considered to be of very limited value, or to actually be detrimental to language acquisition.

Long (1991) proposed a third option, *focus on form*, which he defined as "an occasional shift of attention to linguistic code features—by the teacher and/or one or more students—triggered by perceived problems with comprehension or production" (Long & Robinson, 1998, p. 23). The provision of implicit negative feedback is one way to cause such a focus on form in meaningful interactions (1998, p. 25). Similarly, in his interaction hypothesis, Long (1996) argued for the effective role of negative feedback provided during meaningful interactions for L2 learning (p. 414).

One form of such negative feedback is the recast, defined as "utterances that repeat a learner's incorrect utterance, making only the changes necessary to produce a correct utterance, without changing the meaning" (Nicholas, Lightbown, & Spada, 2001, pp. 732-733). The following is an example of a recast from Sakai (2004).

Learner: *His mother angry.*
Researcher: *His mother is angry?* [recast]

The learner omitted the copula *is* in her production. The recast provided in the next turn by the researcher was a reformulation of the preceding utterance with a correction (i.e., the insertion of the copula) while maintaining the meaning that the learner had intended to express.

One characteristic of recasts is that recasts have dual roles as they provide positive as well as negative evidence. They are defined respectively as “models of what is grammatical and acceptable (not necessarily the same) in the L2” and “information about what is ungrammatical” (Long, 1996, p. 413). In the above example, the recast (*His mother is angry?*) shows what should be uttered (i.e., *is*), and at the same time, it informs that the omission of the copula is not grammatical. Therefore, discourse structures including recasts contain complex information for L2 learners (Gass & Mackey, 2007, p. 185).

Another characteristic of recasts is the implicitness of the provision of negative evidence (Long, 2007; Long, Inagaki, & Ortega, 1998). As a result of this characteristic, the interaction flow is not interrupted at the moment when error correction is called for. Thus, if the effectiveness of recasts is empirically supported, this form of feedback will be useful in communication-oriented L2 classrooms in which the main focus is not on form, but on meaning. The present study is designed to examine the effectiveness of one type of recasts, that is, full recasts (a reformulation of all of an incorrect utterance).

Statement of the Problems

For the past few decades, a large number of studies have been carried out investigating issues related to recasts (e.g., Mackey, 2007, 2012). Most researchers have investigated how recasts are produced in various types of classrooms (e.g., Doughty, 1994, Lyster, 1998a, 1998b; Lyster & Ranta, 1997; Roberts, 1995) and in conversational interactions (e.g., Oliver, 1995; Van den Branden, 1997), or whether recasts are effective for L2 acquisition in classroom or instructional contexts (e.g., Doughty & Varela, 1998; Han, 2002; Ishida, 2004; Mackey & Oliver, 2002) and in laboratory settings (e.g., Iwashita, 2003; Long, Inagaki, & Ortega, 1998; Mackey, 1995, 1999; for a meta-analysis, see Mackey & Goo, 2007). Other researchers have investigated the question of how L2 learners perceive recasts by examining their immediate responses to recasts (e.g., Doughty, 1994; Lyster, 1998a, Mackey & Philp, 1998) and by using stimulated recall (e.g., Egi, 2004, 2007a, 2007b; Mackey, Gass, & McDonough, 2000; Sakai, 2004; see Gass & Mackey, 2000 for a review of the methodological issues related to stimulated recalls).

The first gap in the literature addressed in this study concerns the lack of research on the nature of recasts as sources of negative evidence, that is, information about ungrammaticality of the L2. Nicholas, Lightbown, and Spada (2001) pointed out that although there is an assumption that recasts can provide negative evidence in the field of L2 acquisition, the assumption is problematic and needs to be empirically examined (p. 734). Nevertheless, even though a decade or more has passed, few researchers have dealt with this issue. Thus, this issue is the first gap that remains in the literature. In this study, I aimed to clarify whether recasts can provide L2 learners with negative evidence or not.

The second gap that the present study attempted to fill is related to theoretical hypotheses regarding the mechanism of recasts for L2 learning. Several hypotheses have been proposed to explain the cognitive mechanisms underlying recasts in first language (L1) and L2 acquisition: Saxton's direct contrast hypothesis in his contrast theory (Saxton, 1997, 2000), the additional input hypothesis as in Long's interaction hypothesis (Long, 1996, 2007) and Gass' (1997) interaction model (see Gass & Mackey, 2007 for a further discussion of the model), and Leeman's (2003) enhanced salience hypothesis. These hypotheses provide different explanations of the cognitive mechanisms underlying recasts, but which of these hypotheses best accounts for empirical data remains unclear. In this study, I attempted to test the above-mentioned hypotheses to obtain clearer understanding of recasts.

Purposes and Significance of the Study

The purpose of the present study was twofold. The first purpose was to examine the effects of recasts on narrowing overgeneralized grammar in the L2. In other words, this study addressed the issue of whether recasts provide negative evidence. It is important to better understand this issue because, as White (1987, 1989a) argued, L2 learners need to receive negative evidence in order to retreat from overgeneralized rules that are formed mainly because of L1 influences and that cannot be eliminated only with positive evidence. Based on White's argument, Long (1996) claimed that negative feedback might be essential for L2 learners to narrow overgeneralized rules (p. 414). Previous researchers investigating the effectiveness of recasts have mainly examined L2 learners' learning of information about what is possible in the L2, and little light has been shed on the acquisition of information about what is not possible in the L2. Thus, the findings of this examination the acquisition of knowledge about ungrammaticality of certain forms in the L2 contribute to the accumulation of knowledge in the field.

The second purpose was to test three major hypotheses of the mechanisms underlying recasts: the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis. This purpose is important because testing the competing hypotheses empirically makes it possible to contribute to the advancement of theory building as to the mechanisms underlying recasts. Long (2007) pointed out that it is important to understand the mechanisms of the effective functions of recasts and search for better explanations of the effectiveness of recasts (p. 113). By testing the proposed hypotheses, my aim was to clarify the mechanisms of how recasts work for L2 learning.

The Audience for the Study

The exploration of recasts provides insights for L2 acquisition researchers and L2 teachers. First, L2 acquisition researchers can benefit from this study's contribution to theory building concerning the nature of recasts, specifically, how recasts help L2 learning processes. Broadly speaking, this study is informative in that it addresses the issue of negative evidence for L2 acquisition.

Second, as the nature of recasts becomes clearer, L2 teachers can utilize recasts more efficiently in their classrooms. The three hypotheses tested in this study concern conditions for the function of recasts. Therefore, on the basis of the findings of this study, L2 teachers can obtain information about desirable and facilitative contexts in which recasts work effectively. Recasts are useful because they do not interfere with the flow of communication in communicative language classrooms.

Delimitations

The target structures in this study were adjective-ordering rules and indirect passives. First, in regards to the adjective-ordering rules, this study examined whether recasts help L2 learners learn that the color-size-noun sequence is ungrammatical in English. Because the color-size-noun sequence is allowed in Japanese, Japanese learners tend to accept or produce the color-size-noun sequence in English too. Second, in regards to indirect passives (a type of passivization), this study investigated if recasts help L2 learners learn that indirect passives are ungrammatical in English. Because indirect passives are permitted in Japanese, Japanese learners tend to accept or produce indirect passives in English too. (Detailed explanations of the two structures are provided in the Target Structures section). Thus, one of the delimitations is that this study dealt with syntactic rules only. The findings of this study might or might not be generalizable to the acquisition of morphological and pragmatic rues.

Another delimitation is that these target rules are irrelevant to the resetting of a Universal Grammar (UG) parameter. For example, Izumi and Lakshmanan (1998) argued that the reason why indirect passives are allowed in Japanese is not attributed to a difference in parameter setting but to a different functional status of the Japanese passive morpheme *–rare* from the English passive auxiliary *be* (p. 66). The general issue of how L2 learners retreat from overgeneralized L1 rules has been the focus of most UG-based studies; thus, the findings of this study might or might not be generalizable to the discussion of the roles of negative evidence for parameter resetting in UG-based theories.

Japanese-speaking university students learning English as a foreign language participated in this study. They represent a particular type of L2 learner in terms of native language, culture, and English proficiency; thus, the findings of this study should be generalized cautiously to L2 learners with other native languages, cultures, and educational experiences.

In addition, because this study was experimental in nature, the participants were provided with recasts in a way that differs from natural conversational interaction, particularly in terms of the frequencies and the target of feedback. In addition, the participants were probably more attentive to linguistic forms in a laboratory than in a classroom (Mackey, 2012). Although in order to generalize the findings of this study to a variety of situations, that is, more ecologically valid contexts such as classrooms, more empirical studies are needed. This study is a first step into the inquiry about the effectiveness of recasts as negative evidence.

Terminology

In this section, I provide definitions of the terms *positive evidence, negative evidence,* and *negative feedback* because different researchers use these terms differently.

First, I follow Long's categorization of input using the terms *positive evidence, negative evidence,* and *negative feedback* (Long & Robinson, 1998, p. 19). Data for L2 acquisition can be divided into positive evidence and negative evidence. Positive evidence consists of input and models that offer information about "what is grammatical and acceptable (not necessarily the same) in the L2" (Long, 1996, p. 413). On the other hand, negative evidence offers "direct or indirect information about what is ungrammatical" (p. 413). Negative evidence can be preemptive (as in the form of grammar rules) or reactive (negative feedback). Thus, negative feedback is "a subset of negative evidence" and is referred to as "information learners receive … to the effect that something they (usually just) said or wrote in the L2 was ungrammatical" (Long, 2007, p. 76). Negative feedback can be explicit (overt error correction) or implicit (communication breakdown or recasts).

Several researchers use these terms differently. For example, Schachter (1991) pointed out the interchangeability of the terms *corrective feedback, negative data* or *negative evidence,* and *negative feedback.* They are used in the pedagogical field of second language teaching/learning, the linguistic field of language acquisition, and the psychological field of concept learning. All refer to "externally provided information..., either (1) that the production or activity of that student, child, or subject was in some way anomalous, unacceptable, or deviant, or (2) that the activity produced had not achieved its goal" (p. 89). Thus, according to Schachter, negative evidence and negative feedback are used interchangeably.

Saxton (1997, 2000) defined negative evidence and negative feedback differently from Long and Schachter. He used the term *negative input* to refer to what Long termed negative evidence. Negative input consists of negative evidence and negative feedback. Negative evidence refers to recasts, which occur "directly contingent on a child's grammatical error, and is characterized by an immediate contrast between the child error and a correct alternative to the error, as supplied by the child's interlocutor" (Saxton, 2000, p. 223). On the other hand, negative feedback "occurs directly contingent on a child grammatical error, and provides a non-specific indication that something is amiss within the preceding child utterance" (p. 223). According to Saxton, negative evidence and negative feedback are mutually exclusive.

Saxton (1997, 2000) used the term *positive input* to refer to the positive evidence that is provided not in the juxtaposition of the learner's erroneous utterance but somewhere else. In other words, positive input refers to non-contingent positive evidence of the target structures.

Leeman (2007) made a distinction between feedback and evidence. The former refers to "a mechanism which provides the learner with information regarding the success or failure of a given process" (p. 112) whereas the latter *evidence* refers to the information about whether the structure or form is acceptable (i.e., positive evidence) or unacceptable (i.e., negative evidence). The feedback regarding the success of the communication process is positive feedback whereas the feedback regarding failure is negative feedback. Following Leeman's distinction, Mackey (2012) stated that "linguistic information (i.e. evidence) is provided to a language learner" (p. 114) through the mechanism (i.e., feedback). Leeman (2007) and Mackey (2012) pointed out that the relationship between feedback and evidence types might not be clear and argued for use of the term *error correction.*

Throughout this study, I use Long's categorization and definitions of input in order to avoid confusion about the terminology, unless otherwise indicated. Thus, I ask the following questions interchangeably: "Do recasts provide L2 learners with negative evidence?" (as in the title of this study), "Do recasts function as negative feedback that contains the information about the ungrammaticality or unacceptability of some forms or structures in L2?," and "Do L2 learners receive negative evidence about L2 through recasts?"

Organization of the Study

The organization of this study is as follows. Chapter 2, Review of the Literature, is divided into four main sections: Theoretical Issues, Empirical Studies, Methodological Issues, and Definitional Issues. This chapter ends with a statement of the gaps in the literature, the purposes of this study, and the research questions that guide this study. In Chapter 3, Methods, I present information concerning the participants, the research designs, target structures in this study, the tasks used for testing and treatment, the procedures, and the analytical methods. In Chapter 4, Results, I show the descriptive statistics obtained by means of the testing tasks and the results of the inferential statistics. In Chapter 5, Discussion, I discuss the findings based on the research questions and present the theoretical and pedagogical implications of the study. Finally, I discuss the limitations of the study, make suggestions for further research, and provide concluding remarks in Chapter 6, Conclusion.

CHAPTER 2

REVIEW OF THE LITERATURE

Theoretical Issues

In the first section, I summarize theoretical arguments regarding the roles for negative evidence in L2 acquisition (White, 1987, 1989a). According to White, L2 learners need negative evidence in order to retreat from overgeneralized rules to the narrower target rules. I then review three major hypotheses that have been proposed to explain the mechanisms underlying recasts in the acquisition processes: Saxton's direct contrast hypothesis (1997, 2000), the additional input hypothesis (Gass, 1997; Gass & Mackey, 2007; Long, 1996, 2007), and Leeman's (2003) enhanced salience hypothesis. Saxton argued that the positional characteristics of recasts that are juxtaposed with learners' erroneous utterances help learners notice the errors. Gass and Long argued that noticing negative evidence might not be sufficient for restructuring interlanguage rules and that L2 learners need input to search for further positive evidence. Leeman hypothesized that recasts are effective not because they provide negative evidence but because they provided input with enhanced salience due to negative evidence.

White's Arguments About Roles for Positive and Negative Evidence

White (1987, 1989a) argued for the necessity of negative evidence if L2 learners are to eliminate L2 overgeneralizations. White (1989a) illustrated three types of incorrect hypotheses that learners might construct in order to explain what type of evidence might be necessary to modify the incorrect hypotheses. She pointed out that there are cases in which negative evidence is needed to modify incorrect hypotheses.

The first type of incorrect hypothesis is what she called under-representation (p. 142). Assume that there are two possible grammars A and B and that grammar A is a subset of grammar B: In other words, structures and forms permitted by grammar A are also allowed in grammar B, whereas there are some structures and forms permitted only by grammar B. When the learner internalizes grammar A, the learner's incorrect hypotheses produced by grammar A are an under-representation of the target grammar B. She argued that in this case, what a leaner needs in order to change grammar A to B is positive evidence (p. 142).

The second type is an incorrect hypothesis (grammar A) that should be replaced by different hypotheses (grammar B), where both grammar A and B do not overlap. White mentioned the learning of irregular past tense forms as an example of the second type. The learner overgeneralizes the regular forms of English past tense to irregular verbs and produces incorrect forms by adding the morpheme *-ed* to irregular verbs. These incorrect forms are eventually replaced by the correct irregular past forms, but replacement takes place only after the provision of positive evidence. According to White (1989a), L1 acquisition researchers therefore proposed a uniqueness principle, which states that "any particular semantic concept will have only one syntactic or morphological realization" (p. 143) in order to prevent learners from thinking that both forms are possible in English. White gave another example of this type of incorrect hypothesis: the head-initial and head-final parameter (p. 166). For example, Japanese learners of English have to switch from the L1 value of head-initial to the L2 value of head-final. In this case, she argued that L2 learners might reset the value of the parameter only with positive evidence.

The third type of incorrect hypothesis is a superset of the target grammar. Assume that there are two possible grammars A and B and that grammar A is a superset of grammar B: In other words, structures and forms permitted by grammar B are also allowed in grammar A, whereas there are some structures and forms permitted only by grammar A. When the learner internalizes grammar A, the learner needs to retreat from overgeneralized rules (grammar A) to narrower target rules (grammar B). In this case, White proposed that learners require negative evidence to learn that sentences produced by overgeneralized rules are not grammatical in the target grammar B.

In L1 acquisition theories, the subset principle has been proposed as a learning mechanism that helps learners to avoid the third type of incorrect hypothesis. The subset principle states, “where one hypothesized language is properly contained within another, the acquisition procedure will always guess the smallest language compatible with the positive evidence so far encountered” (Berent, 1994, p. 18). Thus, negative evidence is irrelevant in L1 acquisition. On the other hand, most researchers have argued that the subset principle is not available in L2 acquisition (e.g., White, 1989a), although a few researchers argued for L2 learners’ use of the subset principle (Berent, 1994). White provided evidence from UG-based studies for the non-use of the subset principle in L2 acquisition (pp. 148-165). In addition, White (1987) stated that L2 learners make overgeneralized hypotheses about the target language because of the influence of their

L1 (p. 105). Therefore, White argued that L2 learners might need negative evidence to reset the superset value to the subset value of a parameter or to retreat from a superset grammar to a subset grammar of the L2 (1989a, p. 168; see also White, 1987, p. 107, for her argument concerning this issue).

Overgeneralization of L1 Rules by L2 Learners

Several researchers have shown that L2 learners overgeneralize L1 rules to the target language and that those overgeneralized rules are difficult to eliminate solely with positive evidence (Izumi & Lakshmanan, 1998; Trahey, 1996; Trahey & White, 1993; White, 1989b, 1991). White (1989b, 1991; Trahey & White, 1993; Trahey, 1996) examined the acquisition of the adverb placement of French speaking learners of English. In English, an adverb cannot appear between the verb and its direct object: **Mary ate quickly her dinner* (White, 1989b, p. 136); while in French, the equivalent word order is allowed: *Marie a mangé rapidement le diner* (p. 137). White (1989b) examined whether L2 learners observe the subset principle, in other words, whether they start with less general grammar, which might not require negative evidence for its approximation to the target grammar. As the target structure, she selected the adjacency condition on case assignment, which states that an overt noun phrase must be assigned a case and that it "must be next to its case assigner" like a verb or preposition (p. 136). She argued that there is a parameter of the adjacency condition on case assignment: [+strict adjacency] and [-strict adjacency]. English is a [+strict adjacency] language and allows nothing to intervene between the noun phrase and its case assigner, whereas

French is a [-strict adjacency] language and allows structures violating the adjacency condition with particular types of adjuncts as well as structures following the adjacency condition. Thus, in English, the word order of SVAO (subject + verb + adverb + object) is not admissible, whereas in French, SVAO is grammatical. She also assumed that the [-strict adjacency] value is a more general grammar than the [+strict adjacency] value because the former encompasses the latter: That is to say, she argued for the superset-subset relationship between the [-strict adjacency] value and [+strict adjacency] value. Her participants were 95 French-speaking learners of English (43 intermediate adults, 29 adolescent learners of 13 years of age on average, and 23 adolescent learners of 15 years of age on average), 120 English-speaking learners of French (25 fifth grade partial immersion students, 12 fifth grade early total immersion students, 16 fifth grade submersion students, 24 sixth grade partial immersion students, 25 sixth grade early total immersion students, and 18 sixth grade submersion students), and native speakers of each language (14 for English and 31 for French). The tests consisted of three types of grammaticality judgment tasks: a paced judgment task, a multiple-choice judgment task, and a comparison task. It was found that French-speaking learners of English judged [+strict adjacency] as grammatical, but their judgments about [-strict adjacency] were less accurate than those about [+strict adjacency]. The results suggested that the subset principle might not have worked for these learners; rather, the French-speaking learners of English might have transferred L1 target structures and might have treated English like French by believing that English allows adjacency violations. On the other hand, the English-speaking learners of French judged [+strict adjacency] as grammatically correct; however, their judgments about [-strict adjacency] were conservative, and they tended to judge [-strict adjacency] as ungrammatical. White suggested that the findings for the English-speaking learners of French might be due to either the subset principle or L1 transfer (p. 153).

White (1991) investigated whether negative evidence in the form of explicit grammar explanation and error correction would lead to a retreat from the overgeneralized rules. The target structures, adverb placement in English, were the same as White (1989b); however, she relied on a different theoretical approach to adverb placement. That is to say, she referred to the verb-raising parameter instead of the adjacency condition on case assignment. The parameter might cover a wider range of linguistic differences between English and French, such as negation and question formation (p. 137). Verbs in French must be raised to INFL (inflectional phrase position) from V, thus resulting in the SVAO order. In English, only *have* and *be* are raised to INFL. Her participants were French-speaking learners of English in grades 5 and 6, divided into two groups: the adverb group (n = 82) and the comparison group (n = 56), which she called the question group. Two posttests were administered, the first one immediately after the treatment and the second one five weeks after the treatment. The fifth grade students took another delayed posttest one year later. The tests consisted of a grammaticality judgment task, preference task, and manipulation task. The adverb group received formal instruction about adverb placement for five hours in the first week and performed activities for two hours in the second week. The question group received a similar treatment about a different rule, question formation. The results showed that, on the pretest, both the adverb group and the question group assumed that English is a verb-raising language like French and that only the adverb group demonstrated that they had learned the prohibition of verb raising on the posttest and delayed posttest. However, the effects of formal instruction disappeared in the second delayed posttest one year after the treatment. Thus, she argued that the provision of negative evidence in the form of explicit formal instruction might have at least short-term effects.

Trahey and White (1993) and Trahey (1996) investigated whether positive evidence would be sufficient for L2 learners to disconfirm overgeneralized rules of adverb placement. Trahey and White's (1993) participants were 54 fifth grade French-speaking learners of English. They received an approximately 10-hour input flood of the target structure over 2 weeks. A pretest, immediate posttest, and delayed posttest 3 weeks after the treatment were administered. The tests consisted of a grammaticality judgment task, preference task, sentence manipulation task, and oral production task. The main results showed that, in all the tasks, the input flood facilitated the learning of the grammaticality of the SAV order, which is ungrammatical in French; however, the participants did not learn the ungrammaticality of the SVAO order, which is grammatical in French, through the enhanced exposure of the input lacking the SVAO order. Thus, they suggested that primary linguistic data (PLD), or positive evidence, might not be sufficient for L2 learners to retreat from overgeneralized rules.

Trahey (1996) conducted a follow-up study with the same participants who had received the input flood of adverb placement in Trahey and White's (1993) study and reported the results of the tests administered after a one-year interval. The results showed that the input-flood participants' acceptance of or preference for the SAV order was maintained and that these participants still accepted or produced the SVAO order. Thus, Trahey argued that these long-term changes might reflect changes in the underlying interlanguage systems and pointed out three possible interpretations of the failure of the input flood as to the SVAO order: (a) the two-week input flood was not sufficient in quantity or duration, (b) the input flood did not contain negative evidence, and (c) positive evidence or negative evidence alone was not sufficient. Regarding the third possibility, Trahey stated that negative evidence in the form of explicit grammar explanation and error correction did not lead to long-term effects on retreating from the overgeneralized SVAO order in English because the participants directed too much attention to the structures, and they were not provided with positive evidence in naturalistic communication (see also White, 1991). She suggested that techniques to promote attention to form with the input flood might have successfully helped learners learning that the SVAO order is ungrammatical in English (p. 136).

Izumi and Lakshmanan (1998) conducted a small-scale study to investigate the effects of the provision of negative evidence in the form of formal instruction on the acquisition of indirect passives by Japanese learners of English. In English, passivization involves (a) positioning the internal argument of the verb in the subject position of the passive sentence, (b) changing the verb form to the past participle, (c) inserting the auxiliary *be*, and (d) suppressing the agent of the verb. The agent can be realized with the preposition *by* as an adjunct in the passive sentence. On the other hand, Japanese not only has a direct passive, which corresponds to the English passive, but also an indirect passive (see the Target Structures section of Chapter 3 for detailed explanations and more examples of indirect passives). The researchers showed the following examples (p. 69):

(1) a. Taro-ga se-no takai hito-ni zibun-no mae-ni suwa-rare-ta.

Taro-NOM height-GEN high person-by self-GEN front-in sit-PASSIVE-PAST

"Taro was adversely affected because a tall person sat in front of him."

Lit. *Taro was sat by a tall person in front of self.

b. John-ga Tom-ni konpiuutaa-o tukawa-(r)are-ta.

John-NOM Tom-by computer-ACC use-PASSIVE-PAST

"John was adversely affected by Tom's using a/the/her/his computer."

Lit. *John was used a/the/her/his computer by Tom.

Noun phrases with the nominative case express a person suffering an adverse experience in the indirect passives, as in (1). Thus, indirect passives are referred to as adversative passives as well. Verbs contain the suffix indicating passivization (*-rare)*, but the internal argument with accusative case does not move to the subject position of the indirect passive as in (1-b). Indirect passives can have intransitive verbs as in (1-a) as well as transitive verbs. The researchers predicted that Japanese learners of English would have difficulty learning that the indirect passive in English is ungrammatical because of overgeneralizing their L1 rules to the L2 and because the positive evidence that contains direct passives only does not provide information about the disallowance of indirect passives in English.

The participants in Izumi and Lakshmanan's study were 15 Japanese learners of English studying at the Center for English as a Second Language at Southern Illinois University at Carbondale in the USA. Izumi and Lakshmanan employed a pretest-posttest-delayed posttest design with a control group. The participants were divided into an experimental group (n = 4) and a control group (n = 11). One week after the pretest, the experimental group received two sessions of 90 minutes of formal instruction about indirect passivization spread over two days, whereas the control group did not receive any formal instruction. The instruction included the provision of explicit negative evidence in the forms of a grammatical explanation containing the presentations of indirect passives and the explanation about the ungrammaticality of the indirect passives. Also the instruction also included the provision of positive evidence in the forms of grammatical example passives and alternative expressions of indirect passives.

Five days after the treatment, the first posttest was administered to both groups. Eight weeks after the treatment, the second posttest was administered to only two participants from the experimental group. The test included a translation test, a picture-cued production task, and a grammaticality judgment task. In the translation test, the participants were asked to translate six Japanese sentences into English: one for the direct passive and five for the indirect passive. In the picture-cued production task, the participants were shown pictures about the event happening to a character Tom and asked to describe what happened to Tom. They were also provided with specific verbs for the events. The number of items in the picture-cued production task was eight (two direct passives and six indirect passives), seven (one direct passive and six indirect passives), and six (six indirect passives) for the pretest, the first posttest, and the delayed posttest, respectively. In the grammaticality judgment task, the participants were asked to judge and correct the grammaticality of 16 sentences (eight direct passives and eight indirect passives).

The main findings of Izumi and Lakshmanan (1998) were as follows. First, the results of the pretest showed that most of the participants produced indirect passives and accepted them as being grammatical. For example, in the translation test, 13 of the 15 participants produced at least one indirect passive. The examples produced by the participants were as follows (p. 81): “Mr. Tanaka was stolen stereo” (= Mr. Tanaka is upset because his stereo set was stolen.), “I was eaten final cake by friend” (= The last piece of case, which I had wanted to eat, was eaten by my friend.), and “Taro was cried by his girlfriend yesterday.” (= Taro did not know what to do/Taro was embarrassed when his girlfriend cried.). The results of the picture-cued production task also indicated that twelve of the participants produced at least one indirect passive. In the grammaticality judgment test, four of the participants successfully rejected all the indirect passives, but most of them judged indirect passives as being grammatical. Thus, the findings suggested that the Japanese learners of English overgeneralized the L1 rules concerning indirect passives to the L2.

Second, the results of the posttest showed that the experimental group improved remarkably; on the other hand, the control group persisted in producing and accepting indirect passives. In the translation posttest, none of the participants in the experimental group used indirect passives on the posttest, resulting in the improvement of the mean error scores from 40% to 0%, whereas the mean error scores did not change for the control group (25.5% to 23.6%). Similarly, in the picture-cued production task, the experimental group performed better on the posttest than on the pretest: The mean error scores were 37.5% on the pretest and 0% on the posttest. In contrast, the control group produced more indirect passives on the posttest than on the pretest: The mean error scores were 18.2% on the pretest and 24.2% on the posttest. The same results were obtained for the grammaticality judgment test. The mean error scores on the pretest and the posttest were 43.8% and 0% for the experimental group and 27.3% and 42.0% for the control group. The results of the delayed posttest with two participants from the experimental group suggested that the effects of the formal instruction lasted even after 8 weeks. In sum, the results showed that the provision of explicit negative evidence as well as positive evidence led to retreat from the overgeneralized passive, that is, the indirect passive.

The studies reviewed above (Izumi & Lakshmanan, 1998; Trahey, 1996; Trahey & White, 1993; White, 1989b, 1991) showed that L2 learners made errors derived from overgeneralization of the L1 rule system and suggested that positive evidence or negative evidence alone might not be sufficient to help learners retreat from overgeneralized rules and acquire target-like ones (Trahey, 1996; Trahey & White, 1993; White, 1989b, 1991), whereas a combination of negative and positive evidence might be effective (Izumi & Lakshmanan, 1998).

Roles of Negative Evidence in L2 Acquisition

Several researchers have argued that UG in L2 learning does not require negative evidence (e.g., Schwartz, 1993). For example, Schwartz distinguished performance stemming from competence or UG from learned linguistic behavior, which derives from learned linguistic knowledge. She argued that for the growth of the former, L2 learners require exposure to positive evidence, what she called primary linguistic data (p. 153). Further, she hypothesized that negative evidence in the form of the propositional content cannot feed into competence (p. 157). However, it is important to note that White's argument about how L2 learners abandon incorrect hypotheses might be applied to L2 learning in general. White (1989a) did not confine herself to UG-based studies and suggested that negative evidence might be necessary if L2 learner's interlanguage grammar constitutes a superset of the L2 as well as in the case where parameter resetting from the superset to the subset value is required (White, 1989a, p. 168). Similarly, Gass (1997) argued for the necessity of negative evidence in L2 acquisition in general, whether the target rules are UG-based knowledge or language specific (p. 42). She pointed out that there are a number of cases in which positive evidence is not sufficient for L2 learners to learn that some structures or rules are ungrammatical and argued that L2 learners need negative evidence (p. 100). Likewise, Izumi and Lakshmanan (1998) showed that a learnability problem arises in L2 areas other than parameter-resetting. They assumed that the parameterization of the passive had not been proposed in the literature and hypothesized that Japanese learners of

English assume that indirect passives are permitted in English as well as in Japanese. White's argument can be extended to L2 acquisition in general.

I have reviewed White's arguments that L2 learners require negative evidence in cases where their interlanguage grammar constitutes a superset grammar of the target language, mainly because of the transfer of their L1. However, she did not clearly specify the cognitive mechanisms that would permit negative evidence to reduce the overgeneralized grammar. In the following three sections, I review three major hypotheses concerning these mechanisms for recasts: the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis.

The Direct Contrast Hypothesis

The first hypothesis is Saxton's (1997, 2000) direct contrast hypothesis. Saxton proposed the contrast theory to account for the mechanisms of negative evidence underlying L1 acquisition. According to the contrast theory, the corrective potentials should be "embodied in specific kinds of adult-child interaction" (1997, p. 145). Thus, Saxton argued that the position of a certain utterance in a discourse is crucial and creates a context in which the child can interpret the utterance to have corrective nature.

The contrast theory consists of two hypotheses: the direct contrast hypothesis and the prompt hypothesis. The former is relevant to the mechanisms of recasts; the latter concerns the mechanisms of what he called negative feedback. In this section, I follow Saxton's definitions for his terms (see the Introduction for details regarding the definitions). First, the direct contrast hypothesis, which is one focus of the present study, states:

> When the child produces an utterance containing an erroneous form, which is responded to immediately with an utterance containing the correct adult alternative to the erroneous form (i.e. when negative evidence is supplied), then the child may perceive the adult form as being in CONTRAST with the equivalent child form. Cognizance of a relevant contrast can then form the basis for perceiving the adult form as a correct alternative to the child form. (1997, p. 155)

Thus, the juxtaposition of the child's erroneous utterance and the adult's correct alternative is an important feature. The direct contrast hypothesis predicts that when not juxtaposed with the child's utterance, the adult's correct alternative, which Saxton called positive input, or non-contingent positive evidence, might not work as negative evidence.

Second, the prompt hypothesis states:

> Negative feedback can prompt the child to attend to an ungrammatical form in a previous utterance, and apprehend it as such, in just those cases where the child has prior knowledge of its ungrammaticality. (2000, p. 228)

Because negative feedback does not contain information about how to correct the error, the child is required to possess prior knowledge about the ungrammaticality of a target form. He argued that negative feedback works only "in cases where the child's memory retrieval system has failed" (2000, p. 228).

Saxton tested and confirmed the direct contrast hypothesis using an experimental design (Saxton, 1997) and the direct contrast hypothesis and the prompt hypothesis using a natural corpus (Saxton, 2000). First, Saxton (1997) conducted an experiment with 36 children of ages of 4;9 to 5;6. He created six novel verbs: three words that required a vowel change only (*pro/prew, neak/noke,* and *jing/jang*) and three words that required a vowel change as well as the *–ed* suffix (*streep/strept, sty/stought,* and *pell/pold*). He employed a within-subjects design in which all the participants learned half of the words under the negative evidence condition and the rest of the words under the positive input condition. The words were randomized for individual participants. The participants were exposed to the novel words in the base and *–ing* forms in three 15-to-20-minute training sessions on three consecutive days in order to learn the meanings of the novel words through puppet shows on video. The testing session was carried out immediately after the third training session on an individual basis. Under the negative evidence condition, the participants watched the video and were asked, "What happened?" The participants were induced to produce overgeneralized past forms by adding the *–ed* suffix. They were then provided with negative evidence (i.e., recasts). Under the positive evidence condition, the participants were given no time to answer the question, "What happened?" Instead, they were provided with positive input. Under both conditions, the participants' immediate responses to negative evidence and positive input were analyzed. The results showed that in response to negative evidence, the participants produced correct utterances 29.6% of the time, produced partly correct forms 13.0% of the time, persisted in making the same errors 30.6% of the time, and made move-ons 26.8% of the time. In contrast, positive input did not induce correct forms or partly correct forms at all, but led to the persistence of the same errors (85.2% of the time) or move-ons (14.8% of the time). Thus, he interpreted his results as supporting evidence for his direct contrast hypothesis.

Second, Saxton (2000) examined the immediate effects of negative evidence and negative feedback on the grammaticality of child speech to test the direct contrast hypothesis and the prompt hypothesis. He reanalyzed the naturalistic conversational data elicited from Eve (1;6 to 2;3) in Brown (1973) in order to test his contrast theory. First, he detected 11 categories of grammatical and syntactic errors in Eve's speech: present progressive (auxiliary), prepositions, plural *–s*, irregular past tense, auxiliary verbs, possessive, noun phrase specifier, 3rd person singular *–s*, be copula, subject omission, and object omission. He then coded adult responses to child errors into three categories: negative evidence, negative feedback, and adult move-ons. In addition, he coded the utterances including target forms that do not follow child errors as positive input and non-error-contingent clarification requests. Further, the child responses to the above adult utterances were categorized into use correct (UC), persist-with-error (PE), and child-move-on (CMO); thus, he attempted to examine whether Eve corrected her errors immediately after the feedback. The main findings were as follows: First, when he analyzed the data elicited after Eve had attained 50% accuracy in each target form, the results showed that negative evidence and negative feedback induced more correct use of the target forms (20.5% and 17.4% respectively) than positive input did (9.5%). More specifically, Eve corrected her errors in response to negative evidence statistically more frequently than in response to positive input; however, positive input and adult move-ons did not differ statistically. Thus, he suggested that what is crucial is not the provision of grammatical forms but the juxtaposition of adult grammatical forms to the child erroneous utterance. In other words, his results supported his direct contrast hypothesis. Similarly, negative feedback was found to elicit Eve's correction of the errors significantly more than positive input and adult move-ons. In addition, he found that no-error-contingent clarification requests did not lead to changes in Eve's utterances. Thus, he suggested that negative feedback, that is, error-contingent clarification requests, might be effective for helping learners learn that the target forms are ungrammatical. The results lent support to the other hypothesis, the prompt hypothesis, in his contrast theory. Moreover, because he did not find favorable results for negative evidence and negative feedback in analyzing the total data, he suggested that children's responsiveness to negative input might depend on their developmental stage in L1 acquisition.

It should be noted that Saxton limited the direct contrast hypothesis to situations in which the learner already knows two forms that fulfill equivalent functions. He stated that "The Direct Contrast hypothesis, on the other hand, deals with a situation in which the child already possesses, say, both *feeled* and *felt* in her productive repertoire, or can readily apprehend their functional equivalence" (1997, p. 157). Thus, learners must have noticed some conflict between the two forms at the moment that they are provided with negative evidence.

The direct contrast hypothesis needs to be examined with a variety of grammatical rules by various measurement methods. First, as Saxton himself admitted, both studies (Saxton, 1997, 2000) examined the participants' immediate responses to negative input. It remains unclear whether the effects are durable. Second, Saxton examined overgeneralized errors in irregular past forms of verbs (1997) and 11 morpho-syntactic errors, such as present progressive, preposition, plural *–s*, subject omission, and object omission (Saxton, 2000). Because incorrect forms of these features and the corresponding target correct forms do not constitute superset-subset relations or at least overlapping relations, these errors might be overcome with positive evidence only, as White (1989a) illustrated as the second type of incorrect hypothesis. Saxton (1997, 2000) did not discuss the case in which learners have two forms or structures that might lead to a learnability problem. Thus, it remains unclear whether the direct hypothesis is applicable to cases in which overgeneralized rules should be reduced to narrower target rules.

The Additional Input Hypothesis

The second hypothesis investigated in this study was proposed as part of the interactionist position. I summarize Long's (1996, 2007) interaction hypothesis, and then Gass' (1997) interaction model (Gass & Mackey, 2007). Both researchers hypothesized that noticing negative evidence might not be sufficient for restructuring interlanguage rules and that L2 learners need input to search for further evidence. Thus, I call their positions *the additional input hypothesis*.

Long (1996) considered negative feedback to play an important role in L2 acquisition. In his interaction hypothesis, he proposed:

> ...environmental contributions to acquisition are mediated by selective attention and the learner's developing L2 processing capacity, and that these resources are brought together most usefully, although not exclusively, during *negotiation for meaning*. Negative feedback obtained during negotiation work or elsewhere may be facilitative of L2 development, at least for vocabulary, morphology, and language-specific syntax, and essential for learning certain specifiable L1-L2 contrasts. (1996, p. 414)

Certain specifiable L1-L2 contrasts refer to L2 rules that cannot be learned only with positive evidence, such as the placement of English adverbs for French learners of English (Trahey, 1996; Trahey & White, 1993; White, 1989b, 1991). More specifically, Long (1996) went on to argue that L2 learners need to notice the absence or ungrammaticality of the overgeneralized rules of L1 to L2 with the help of negative evidence such as grammar rules, explicit error correction, recasts, and communication breakdowns followed by repair sequences (p. 428). Thus, recasts are one form of negative evidence that might increase the likelihood that L2 learners notice otherwise problematic forms. Later, Long (2007) claimed that recasts have several psycholinguistic advantages for L2 learning (p. 114). For example, because the L2 learner already understand at least part of the message of a recast provided immediately after the learner's own utterance, it is easier for the learner to direct attentional resources to the form of the recast and to make a form-function mapping. Also, he argued that because the recast is provided based on the learner's own message, the learner is more motivated to attend to the recast. Lastly, the juxtapositioning of the learner' utterance and the recast makes it possible for the learner to notice the contrast of the two.

However, Long did not consider noticing non-salient forms to be sufficient for internalization. Regarding the mechanisms of recasts, Long (1996) stated:

> A mechanism is posited whereby, while correct form-meaning associations are strengthened both by positive evidence *and* that negative feedback that contains positive evidence, incorrect associations are weakened and in some cases ultimately relinquished altogether as a result both of negative evidence *and* prolonged absence of support in the input. (p. 430)

According to this citation, he hypothesized that after receiving negative evidence, L2 learners require further input to search for confirmation of the target hypotheses.

Gass (1997) also argued that L2 learners not only need to receive negative evidence but also need exposure to additional input in order to obtain confirmatory or nonconfirmatory evidence (p. 144). More specifically, Gass and Mackey (2007) considered interaction to be important because L2 learners can receive information about both positive and negative evidence in the interaction (p. 178) and stated once L2 learners notice their errors in production, their next task is to identify the problem and to determine how to change their interlanguage grammar (p. 179). Therefore, Gass and Mackey (2007) argued that learners require further input to search for more evidence confirming or disproving their hypothesis.

Long and Gass differed in that Long emphasized that negative evidence might increase the salience and noticeability of forms, whereas Gass emphasized L2 learners' search for further evidence in the input after negative evidence; however, both considered noticing of errors through negative evidence and search for further evidence in the input to be necessary.

The Enhanced Salience Hypothesis

The third hypothesis investigated in this study is Leeman's hypothesis, which I call *the enhanced salience hypothesis*. She hypothesized that recasts are effective not because they provide negative evidence but because they provide input with enhanced salience due to negative evidence so that L2 learners will have greater chances to attend to the target forms and to incorporate the relevant rules into the interlanguage grammar (2003, p. 45).

Leeman (2003) investigated whether enhanced salience alone promotes L2 development or whether negative evidence is also necessary for L2 development. Although recasts and models (i.e., positive evidence) have been investigated in previous empirical research, enhanced salience was not taken into consideration. Therefore, Leeman compared the effectiveness of four types of input (recasts, negative evidence, enhanced salience, and unenhanced input) on the learning of noun-adjective agreement in Spanish in order to investigate which aspect of recasting would lead to L2 acquisition: positive or negative evidence.

The participants, 74 university students studying in the United States, were randomly assigned to one of the four conditions. Noun-adjective agreement in gender and in number in Spanish was selected as the target structures because these morphological features have low perceptual salience. A pretest, posttest, and delayed-posttest design was employed. On the first day, the participants performed several distractor tasks and then took the pretest, treatment tasks, and the posttest in a one-hour session. Approximately one week later, they carried out a distractor task, took the delayed posttest, and completed a debriefing questionnaire. The three tests were all oral production tasks in which the participants were asked to find and describe as many differences between two pictures as possible. Each task included 32 items designed to elicit the target structures. Gender and number of nouns were balanced.

In the treatment tasks, each lasting approximately 20 minutes, the participants individually performed two communicative tasks with the researcher in four conditions. In the first task (an object placement task), the participants gave directions to the researcher while in the second task (a catalog-shopping task), the researcher gave directions to the participants. Thus, the treatment session had two phrases: production and comprehension.

In the recast group, the researcher provided recasts of ungrammatical utterances containing noun-adjective agreement errors in the first task. In the negative evidence group, participants were informed of the original utterance implicitly, but were not provided with correct forms. Although Leeman did not explain the techniques used in the negative evidence group, the example (p. 49) indicates that repetition of the erroneous utterances was used. In the enhanced salience input group, the participants were not provided any feedback on the target structures. In the second task, they were exposed to the researcher's directions, which contained positive evidence whose salience was enhanced by stress and intonation. In the unenhanced input group, which served as a control, the participants did not receive any feedback on the target structure in the first task and were provided with directions containing positive evidence without any stress in the second task. The number of opportunities for production and the quantity of positive evidence that the participants received were controlled through the two tasks and across the four groups.

Leeman focused on feminine and plural nouns and calculated the ratios of target agreement to obligatory contexts separately for gender and number. The results showed that, for both gender and number agreement, there were statistically significant differences among the groups both on the posttest and the delayed posttest. More specifically, the recast group and the enhanced salience input group performed better than the control group for both agreements on the posttest; however, the superiority of the recast group over the control group disappeared on the delayed test for gender agreement. No statistically significant differences were found between the negative evidence group and the control group or between the three treatment groups (recast, negative evidence, and enhanced salience). Thus, it was found that both recasts and enhanced salience were effective; but exposure to negative evidence was not (p. 54).

Leeman interpreted the findings as evidence supporting the benefits of recasts even for the acquisition of non-salient forms. She also suggested that the enhanced salience of the input in recasts might be more crucial for L2 development than the negative evidence provided in the recasts. Because there were no statistically significant differences among the three groups (recast, negative evidence, and enhanced salience), she pointed out that her results did not suggest that negative evidence might not lead to L2 acquisition. However, she argued that an account that the recasts are effective because they provide negative evidence might not be plausible, on the basis of the fact that the negative evidence group did not outperformed the control group (p. 56).

Leeman's suggestion about the role of recasts might apply only to the acquisition of linguistic features that might not otherwise be noticed because of a lack of salience. As White (1987, 1989a) pointed out, because L2 acquisition involves overgeneralization about L2 rules, perhaps due to L1 transfer, positive evidence would not be sufficient for L2 learners to notice that the overgeneralized rules are not permitted in the target language. In order to examine Leeman's claim that negative evidence might be unnecessary, researchers need to examine the effects of recasts on narrowing overgeneralized rules.

Summary

In this section, I summarize the similarities and differences among the three hypotheses.

Both the direct contrast hypothesis and the additional input hypothesis state that the juxtaposition of the learner utterance and the recast makes the recast function as negative evidence. The two hypotheses predict that recasts provide negative evidence, whereas non-contingent positive evidence does not work as negative evidence.

The direct contrast hypothesis and the additional input hypothesis might differ in the sense that the latter states that further input after the provision of negative evidence (in this case, in the form of recasts) is necessary for learners to modify their overgeneralized hypotheses. Thus, the direct contrast hypothesis predicts that the provision of recasts might be sufficient to bring about changes in the learner's interlanguage; on the other hand, the additional input hypothesis predicts that the learning effects might differ between L2 learners who receive additional input after recasts and those who do not have additional input after recasts.

According to the enhanced salience hypothesis, the juxtaposition of the recasts helps increase the salience of the input and provide negative evidence; however, it posits that what plays an important role for L2 acquisition is not the provision of negative evidence but the enhanced salience of positive input. Thus, the enhanced salience hypothesis predicts that L2 learners benefit from receiving positive evidence with enhanced salience to the same degree as those who receive recasts. As reviewed above, Leeman (2003) found supporting evidence for the enhanced salience hypothesis in the learning of adjective-noun agreement. Thus, it remains unsolved whether the enhanced salience hypothesis holds for the case in which L2 learners learn that some forms are ungrammatical. If Leeman's hypothesis is applicable to the aspect of ungrammaticality, it predicts that enhanced salience is as effective as recasts in narrowing overgeneralized grammar.

To sum up, the variables in the three hypotheses are juxtaposition, additional input, and enhanced salience.

Empirical Studies

In this section, I review previous studies that addressed the issue of whether recasts are effective in the learning that some forms are ungrammatical in the target language or whether recasts lead to noticing of ungrammaticality of certain forms. I discuss these studies mainly in terms of the target structures (i.e., which types of incorrect hypotheses in White's illustrations these studies treated) and possibly relevant hypotheses, although the relevance to any of the above three hypotheses was not clearly stated in any of the studies.

I divide the studies according to the measurement methods. I first review studies using grammaticality judgment tests and/or elicited imitation tests and then studies in which the retreat from ungrammatical rules was measured indirectly by examining decreases in interlanguage forms or structures. Third, I review the studies in which the noticing of ungrammaticality was investigated through stimulated recall.

Studies Using Grammaticality Judgment Tests and/or Elicited Imitation Tests

Several researchers (Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Loewen & Nabei, 2007; see Ayoun, 2001, 2004, for a study of written feedback) used grammaticality judgment tests and/or elicited imitation tests. For the purpose of the present study, the question in mind is whether these studies provided information about the effectiveness of recasts in terms of learning ungrammaticality of erroneous forms or structures, as measured by the tests using ungrammatical sentences.

Ellis, Loewen, and Erlam (2006) examined the effects of implicit feedback (recasts) and explicit feedback (metalinguistic feedback) on the acquisition of regular past tense *–ed*. Ellis, Loewen, and Erlam operationalized metalinguistic feedback as involving (a) repetition of an error and (b) metalinguistic information about the relevant rules, but not containing the correct forms (p. 350). Thus, they considered metalinguistic feedback to provide only negative evidence such as “You need past tense” (p. 341), not positive evidence. Furthermore, the explicitness of metalinguistic feedback was increased with the repetition of an error. On the other hand, the recasts provided in their study were “partial recasts of those portions of learners’ utterances that contained an error” (p. 350).

The participants were 34 learners of English studying at a private language school in New Zealand. Three intact classes were assigned to three conditions: recasts (n = 12), metalinguistic feedback (n = 12), and control (n = 10). They employed a pretest-immediate posttest-delayed posttest design with a control group. The testing instruments were an oral elicited imitation test with 36 sentences (including six grammatical and six ungrammatical sentences for the regular past form and the comparative form, and 12 distracters) as a measurement of implicit knowledge and two measures of explicit knowledge, an untimed grammaticality judgment test with 45 sentences (including seven grammatical and eight ungrammatical sentences for the regular past form and the comparative form, and 15 distractors) and a meta-language test. A pretest was administered five days prior to the treatment sessions. The treatment sessions were conducted for two consecutive days and lasted one hour in total. On both days, the participants were asked to carry out the story-retelling tasks with picture sequences in triads, which had been designed to elicit the use of past tense. During the tasks, one of the researchers provided recasts to one group and metalinguistic feedback to another. Thus, the feedback was given to the participants in a classroom situation. The control group did not perform these tasks but received normal instruction that did not target the past tense forms. The immediate posttest was administered one day after the treatment, and the delayed posttest was completed 12 days later.

Their main results were as follows. First, at the beginning of the study, the participants possessed limited implicit knowledge of the regular past tense forms as indicated by the low scores on the oral imitation pretest, although they showed high levels of explicit knowledge of the target structure as indicated by the high scores on the grammaticality judgment pretest, in particular, in the judgment of ungrammatical sentences. Second, corrective feedback (both recasts and metalinguistic feedback) led the participants in the experimental conditions to get higher scores on the oral imitation delayed posttest than the participants in the control group. Thus, the researchers argued that both types of feedback were beneficial for the learning of implicit knowledge. Third, they found that metalinguistic feedback was more effective than recasts on all three types of tests, and particularly with new items that had not been dealt with in the treatment sessions; thus, they suggested that system learning, not item learning, might have been brought about by explicit feedback.

Ellis (2007) investigated the differential effects of explicit and implicit feedback on two target structures: the regular past tense *–ed* and the comparative *–er*. The data about the regular past tense were reported separately in Ellis, Loewen, and Erlam (2006); thus, the participants and the measurement instruments were the same as described above. The participants who received recasts for the past tense, that is the recast group in Ellis, Loewen, and Erlam (2006), received metalinguistic feedback for the comparative form; on the other hand, the participants who were provided with metalinguistic feedback for the past tense, that is, the metalinguistic group in Ellis, Loewen, and Erlam, were given recasts for the comparative form. Thus, the group assignment to the feedback conditions was counterbalanced. The recast group received 42 recasts for the past tense and 22 recasts for the comparative; the metalinguistic feedback group received 25 instances of feedback for the past tense and 19 for the comparative.

Ellis did not find statistically significant interaction effects of the structures, testing times, and the groups (i.e., feedback types) in the oral imitation test for the grammatical sentences and in the grammaticality judgment tests for the grammatical and ungrammatical sentences, but he found a statistically significant interaction effect of the three variables in the oral imitation test for the ungrammatical sentences. Specifically, in the oral imitation test with the ungrammatical sentences, the participants provided with recasts showed similar improvement for the two target structures. The mean accuracy rate for the past tense improved from .194 on the pretest to .319 on the immediate posttest to .375 on the delayed posttest for the past tense; and similarly, the mean accuracy rate for the comparative form improved from .264 on the pretest to .472 on the immediate posttest to .639 on the delayed posttest for the comparative form. On the other hand, the participants provided with metalinguistic recasts improved sharply from the pretest to the immediate posttest (.140 to .583) and made a moderate improvement from the immediate posttest to the delayed posttest (.583 to .639) for the comparative form, whereas the scores of the past tense form improved slightly from the pretest to the immediate posttest (.333 to .375) and rose sharply from the immediate posttest to the delayed test (.375 to .633). He also reported that the main effects for group were not significant for grammatical sentences in the oral imitation test and ungrammatical and grammatical sentences in the grammaticality judgment test. The main effect of group was significant for ungrammatical sentence in the oral imitation test; the post hoc analyses showed that the metalinguistic group performed better than the control group and that no significant difference was found between the recast group and the control group. Based on these results, he stated that "only the metalinguistic feedback was found to be effective in promoting acquisition of both structures" (p. 357); however, it remained unclear whether the improvement of the recast group was significantly different from that of the control group or that of the metalinguistic feedback group because he did not report the results of the interaction effect of group and time.

He discussed his findings concerning the differential effects of metalinguistic feedback on the two structures in terms of (a) the extent of well-formedness of the learners' explicit knowledge and (b) the frequency of exposure to the target forms the learners had. His participants had less explicit knowledge of the comparative form than of the past tense form. The explicit feedback might have facilitated external scaffolded attention and internally motivated attention, leading to explicit knowledge and explicit memory. Therefore, metalinguistic feedback showed an immediate effect on the participants' performance on the oral imitation test with ungrammatical sentences, which might have tapped explicit knowledge rather than implicit knowledge. On the other hand, the metalinguistic feedback might have helped the participants recall knowledge that they already had in the case of the past tense form. Ellis stated that recalling the explicit knowledge might have enabled the participants to be more attentive to the past tense form in the input that they received outside the class between the immediate posttest and the delayed posttest and speculated that because the input probably contained a great number of past forms, implicit learning occurred on the delayed posttest (p. 359).

Loewen and Nabei (2007) examined comparable effectiveness of three different feedback types (clarification requests, recasts, and metalinguistic feedback) by using timed and untimed grammaticality judgment tests and oral production test. Their participants were Japanese university students learning English, who were in two intact university classes. The participants in one class formed groups of four, each of which was assigned to one of the four conditions: clarification requests (two groups, n = 8), recasts (three groups, n = 10), metalinguistic feedback (two groups, n = 7), and no feedback (three groups, n = 10); the other class (n = 31) constituted the control group. The target structure was English question formation. Their study was a pretest-posttest quasi-experimental design using two intact classes. For the pretest, the participants took timed and untimed grammaticality judgment tests and an oral production task; they were provided with the treatment during two communicative activities (a spot-the-difference task and a guess-the-storyline task) lasting for 30 minutes one day after the pretest; and one week later they took the three tests for the posttest. The timed and untimed grammaticality judgment tests consisted of 40 sentences: 24 target sentences (12 grammatical sentences and 12 ungrammatical sentences) and 16 distractors. The two tests used the same sentences, but they differed in the time for answering: the participants were asked to respond in 1.8 to 5.0 seconds for the timed grammaticality judgment test whereas the time limit was not set for the untimed grammaticality judgment test. Loewen and Nabei reported on the basis of the transcripts of the treatment interactions that the recast groups received 18.7 recasts and 0.3 clarification requests on average, the clarification request groups received 18 clarification requests on average, the metalinguistic feedback groups received 5.6 instances of metalinguistic feedback and 4 recasts on average, and no feedback groups received almost no feedback.

First, an analysis of the performance on the 24 sentences (combining grammatical and ungrammatical sentences) of the untimed grammaticality judgment showed that for all the conditions, no increase was observed from the pretest to the posttest and that no statistically significant differences were detected among the groups. Furthermore, the researchers examined the grammatical and ungrammatical sentences separately. The interaction effect of the test (pre and posttests) and grammaticality was significant, showing that the scores for the grammatical sentences increased while the scores for the ungrammatical sentences decreased from the pretest to the posttest. Because there was a statistically significant main effect of grammaticality, it was shown that the performance on the ungrammatical sentences was better than on the grammatical sentences.

Second, the researchers reported the results of the timed grammaticality judgment tests. On the combined data of the grammatical and ungrammatical sentences, the main effects of test and feedback group were statistically significant, and the interaction effect of test and feedback group was statistically significant. A post hoc analysis of the gain scores from the pretest to the posttest showed that the feedback groups (clarification requests, recasts, and metalinguistic feedback) improved more than the no feedback group and the control group. In addition, the researchers analyzed the grammatical and ungrammatical sentences separately. Overall, the scores for the grammatical sentences were higher than for the ungrammatical sentences, and the difference was statistically significant. On the other hand, the interaction effects of grammaticality and feedback group and of test and grammaticality were not significant.

Finally, the researchers reported the analysis of the performance on the oral production task. They reported the frequencies of questions in each developmental stage of question formation. The results did not show any clear indications of improvement in question formation from the pretest to the posttest.

Here, I discuss the three studies. It is important to note that Ellis, Loewen, and Erlam (2006) as well as Ellis (2007) did not attempt to examine whether corrective feedback affects the learning about the ungrammaticality of erroneous sentences in target structures, that is, the interlanguage use of the present form for the past reference or simple form for the comparative adjective. Instead, their purpose was to investigate "whether it [corrective feedback] enables learners to gain greater control over a structure they have already partially mastered" (Ellis, Loewen, & Erlam, 2006, p. 351). On the untimed grammaticality judgment pretest with ungrammatical sentences, the scores of the three groups were 85.4%, 84.4%, and 78.8%, respectively. This result indicates that the participants had already learned the ungrammaticality of the use of present forms for the past reference; thus, their focus was on the effects of corrective feedback on the acquisition of implicit knowledge, which they defined as "knowledge that learners are only intuitively aware of and that is easily accessible through automatic processing" (p. 340); that is, whether corrective feedback is effective for developing increased automaticity in the access of the relevant knowledge. The similar point can be made for the results of the comparative form (Ellis, 2007).

It is also important to point out that the target structures of these three studies might not concern the third type of incorrect hypothesis proposed by White (1989a). First, the errors as to the regular past tense *–ed* and the comparative degree *–er*, which Ellis (2007) and Ellis, Loewen, and Erlam (2006) targeted, can be categorized as the second type of incorrect hypothesis discussed by White (1989a). That is to say, the incorrect hypothesis (in this case, no attachment of a past or comparative marker) and the correct hypothesis (in this case, addition of *–ed* to a verb or *–er* to an adjective) do not overlap each other. Learners can replace the incorrect hypothesis by the correct one only with the help of positive evidence. The question formation targeted by Loewen and

Nabei (2007) as well might be related to the first type of incorrect hypothesis (White, 1989a): that is, the incorrect hypothesis constitutes a subset of the target hypothesis. Loewen and Nabei used processability theory (Pienemann, 1998a) as a theoretical framework of L2 development of question formation. According to processability theory, developmental stages in L2 can be explained by acquiring processing procedures (word/lemma access, the category procedure, the phrasal procedure, the S-procedure, and the subordinate-clause procedure) accumulatively. For example, question formation of stage 1 requires acquisition of word/lemma access; question formation of stage 2 requires acquisition of both word/lemma access and the category procedure. Thus, the processing procedures necessary for the lower-stage question formation constitute a subset of the processing procedures necessary for higher-stage question formation (see Sakai, 2008 for empirical data with Japanese university students that showed the accumulative nature of different levels of question formation). To sum up, these studies did not examine whether recasts provide learners with negative evidence so that learners can retreat from overgeneralized rules.

Another point is the treatment used in Ellis, Loewen, and Erlam's (2006), Ellis's (2007), and Loewen and Nabei's (2007) studies. Because their studies were carried out in a classroom situation in which the opportunity to produce and listen to the target structures was not controlled, it is not clear whether the effects of corrective feedback can be attributed to cognitive comparison in the juxtaposition of the erroneous utterance and feedback, more salient positive input, or the provision of positive evidence in addition to negative evidence. Ellis, Loewen, and Erlam showed the following example for the recast condition (2006, p. 353):

Learner:	… they saw and they follow follow follow him
Researcher:	Followed
Learner:	Followed him and attacked him.

This learner had the opportunity to hear the target form *followed* once from one of the researchers and had the opportunity to make an error in the first utterance and to imitate the correct form in the second utterance. On the other hand, the other participants in the room had a chance to hear the correct forms twice (once from the researcher and once from the learner) and had no chances to produce the target forms. In terms of the direct contrast hypothesis, the researcher's recast, *Followed,* might have assisted cognitive comparison of the error and the correct form only for the learner at whom the recast was directed. The listeners were assigned a greater cognitive burden than the learner because they had to understand the learner's utterance and process the feedback in terms of grammaticality at the same time; on the other hand, as Long (2007, p. 78) argued, the learner was able to allocate more attentional resources to the form in the feedback and probably made successful form-function mapping because he or she had already understood the meaning of his or her own utterance. At the same time, the listeners might have received the positive evidence that became more salient by repetitive exposure. Ellis, Loewen, and Erlam gave the following example for the metalinguistic feedback condition (p. 353):

Learner:	He kiss her
Researcher:	Kiss—you need past tense.
Learner:	He kissed.

In this example, the listeners (i.e., everyone except for the learner in the excerpt) received negative evidence from one of the researchers and additional positive input from the learner. Therefore, although the Ellis (2007), Ellis, Loewen, and Erlam (2006), and Loewen and Nabei (2007) studies were unique and interesting in that they employed classroom activities as the treatment tasks and improved external validity, it still remains unclear how recasts or other feedback facilitates L2 learning. In other words, it is difficult to use the findings of these studies to examine the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis concerning the mechanisms of recasts for L2 learning.

Studies Reporting Decreases in Interlanguage Forms

In this section, studies in which interlanguage analysis was utilized (Doughty & Varela, 1998; Ishida, 2004; Loewen & Nabei, 2007) are reviewed. For the purpose of the current study, the question is whether recasts help decrease interlanguage forms in L2 learners' production.

Doughty and Varela (1998) investigated the effects of an incidental focus on form in content-based science classes. They used two intact ESL classes; the participants were assigned to the FonF (focus on form) group (n = 21) and the control group (n = 13), respectively. They employed a pretest-posttest-delayed posttest design with a control group. Both groups received two-week science lessons, in which the participants in the FonF group received what they called *corrective recasting* consisting of repetitions and recasts (pp. 123-124), whereas the participants in the control group did not receive such intervention. The target structure was past and conditional time reference. The participants' production of the target structures was elicited through written reports and oral presentations. Their performance was coded into three categories: target-like encoding (TL), interlanguage encoding (IL), and non-target-like encoding (NTL). The first category referred to the case in which the participants encoded past or conditional time reference in target-like ways; the second category referred to attempts at producing past or conditional time reference in non-target-like forms; the last category referred to non-marking for past or conditional.

The FonF group improved the medians of the TL and IL categories from the pretest to the posttest to a statistically significant degree on both the written and oral tests. The gains on the TL and IL categories were maintained on the oral delayed posttest; on the other hand, only the gain on the IL category was maintained on the written delayed posttest, but the gain on the TL category was not. The medians of NTL decreased from the pretest to the posttest and the losses shown on the delayed posttest were maintained on both the written and oral tests. In contrast, the control group, in general, showed no changes from the pretest to the delayed posttest on both the oral and written tests; one exception was an increase from pretest to posttest on the written test although the increase disappeared on the delayed posttest. Thus, the implementation of incidental focus on form in the content-based lessons was successful. Based on their results and their observations of the lessons, the researchers discussed the feasibility of implementing focus on form tasks in communicative classrooms.

In a time-series design, Ishida (2004) carried out four case studies in concurrent replication to investigate the effects of recasts on the learning of the aspectual form *–te i-(ru)* in Japanese. Her participants were four university students studying Japanese at a university in the United States. The participants individually took part in sessions held twice a week for four weeks in which they joined conversations with the researcher. The pretests were administered in the first two sessions; and the posttests in the last two sessions. The remaining four sessions (the third to the sixth sessions) were treatment sessions, during which the researcher provided the participants with recasts whenever she felt it was necessary to confirm what the participants intended to say. Two of the participants took part in another session, that is, a delayed posttest, seven weeks after the eighth session. Accuracy was measured through a target-like use (TLU) analysis. The TLU index was calculated by transforming into a percentage the value obtained by dividing the number of correct uses by the combined number of obligatory contexts and overuses. The results showed that the mean TLU was low during the pretest sessions, increased sharply during the treatment sessions, and was maintained during the posttest sessions. It was also found that the TLU increase was correlated to the number of recasts. Ishida also analyzed the use of the aspectual form in Japanese in terms of different lexical aspect classes and different aspect meanings in order to test whether the Japanese language data obtained from her participants would conform to the aspect hypothesis. An examination of the delayed posttest for two of the participants showed that their high TLUs were maintained and suggested that the effects of recasts might be durable.

Loewen and Nabei (2007), whose study was mentioned in the previous section, also analyzed performance on the oral production task. They coded the questions according to the developmental stages of question formation. According the developmental stages on which their analysis was based (Pienemann, 1998a, 1998b), Stage 3 refers to WH-Fronting (as in *What Ken read yesterday?*) or Do-Fronting (e.g., *Do John is a teacher?*). In most instances, questions coded as Stage 3 are ungrammatical (except for such interrogatives as *Who is running?* and *Do you like baseball?*). Thus, a decrease in the use of Stage 3 could be used as evidence indicating that a learner might avoid ungrammatical sentences. In Loewen and Nabei, such a decrease was not observed from the pretest to the posttest (p. 372): 1 to 3 for the clarification request groups, 2 to 11 for the recast group, and 4 to 4 for the metalinguistic feedback groups. Unfortunately, the researchers did not provide the frequencies for individual participants or show the questions produced by each participant, and thus it is not possible to discuss their results further here.

Interlanguage analysis was utilized in Doughty and Varela (1998) and Ishida (2004), but on the basis of their results, it is not clear whether recasts effectively helped the learners to learn the ungrammaticality of interlanguage forms. Doughty and Varela's results showed an increase rather than a decrease of interlanguage forms. In their study, the participants who had received corrective recasts decreased non-target-like use; that is, "the verbs were either not marked for any time meaning or prediction meaning or were marked inappropriately, such as in the cases of using present marking for past reference or future marking for past conditional reference" (p. 128). Thus, Doughty and

Varela's results showed that corrective recasts were effective for learners to learn the past or conditional time reference, but did not show that they were effective for reducing interlanguage forms. Ishida used the TLU index and showed that her participants improved their TLUs. The improvement of the TLU index suggested that the non-target-like use and overuse of the target form decreased. However, the decreases in interlanguage forms were not directly shown in her study.

It is difficult to attribute the effects of the treatment to juxtaposition of the recasts, additional input, or the enhanced salience in Doughty and Varela (1998), Ishida (2007), and Loewen and Nabei (2007). Doughty and Varela's (1998) study and Loewen and Nabei (2007) were carried out in classroom situations. The former study focused on corrective recasts, that is, repetition of errors plus recasts. Ishida's (2004) study was longitudinal in nature. Thus, as pointed out for Ellis (2007), Ellis, Loewen, and Erlam (2006), and Loewen and Nabei (2007), it is difficult to isolate the variables of juxtaposition of the recasts, additional input, and the enhanced salience in these studies.

Studies Using Introspective Methods

In this section, research using introspective methods (Egi, 2007a; Gass & Lewis, 2007; Mackey, Gass, & McDonough, 2000; Nabei & Swain, 2002; Roberts, 1995; Sakai, 2004, 2011) is reviewed. I discuss previous studies in terms of whether they attempted to investigate whether L2 learners notice the ungrammaticality of certain forms.

Early researchers (e.g., Nabei & Swain, 2002; Roberts, 1995) investigated whether L2 learners notice the existence of feedback provided in the interaction. For example, Nabei and Swain carried out a case study with one L2 learner of English. Each week, they observed and videotaped a class, administered a grammaticality judgment test, and had a stimulated recall interview with the learner. This cycle was repeated for 10 weeks. The codes for the analysis of the stimulated recall interview were (a) attending to meaning, (b) attending to language, and (c) noticing feedback. The third category was devised to analyze the learner's "explicit recognition of the teacher providing feedback" (p. 51). The researchers introduced excerpts from the stimulated recall interviews and analyzed them qualitatively and reported that the learner noticed feedback provided by her teacher.

Several researchers (Gass & Lewis, 2007; Mackey, Gass, & McDonough, 2000) went on to investigate how L2 learners interpret recasts. More specifically, these researchers examined whether L2 learners recognize the corrective intention of the feedback accurately. For example, Mackey, Gass, and McDonough investigated the perception of interactional feedback including recasts by 10 learners of English as a second language and seven learners of Italian as a foreign language. Each participant carried out a communicative task with a native (or near-native) speaker in which he or she received interactional feedback including recasts as to his or her incorrect utterances. After the task, the participant watched a videotape of the task and was asked to report his or her thoughts at the time of the task. Mackey, Gass, and McDonough analyzed the interactional feedback episodes during the communicative tasks and the stimulated-recall comments. Their codes were (a) phonological episode, (b) morpho-syntactic episode, (c) lexical episode, and (d) semantic episode for the interactional feedback episodes, and (a) lexis, (b) semantics, (c) phonology, (d) morpho-syntax, (e) no content, and (f) unclassifiable for the stimulated-recall comments. The researchers found that their participants perceived feedback about lexis relatively accurately whereas the accuracy of their correct perception morpho-syntactic feedback was limited. The results for the perception of feedback about phonology varied among English learners and Italian learners: English learners recognized phonological feedback being as such accurately, whereas the Italian learners' perception of phonological feedback was not very accurate. They further analyzed the feedback types and reported that 75% of the recasts followed morpho-syntactic feedback; 14% followed phonological feedback; and 11% followed lexical feedback. Thus, the researchers argued that although recasts were a frequent form of corrective feedback on morpho-syntactic errors, recasts for morpho-syntactic errors were not an effective way to promote learners' accurate perceptions about the feedback (p. 493). Gass and Lewis (2007), who also used their coding system, examined perceptions of feedback by 13 heritage and non-heritage learners of Italian.

In the studies mentioned above (Gass & Lewis, 2007; Mackey, Gass, & McDonough, 2000; Nabei & Swain, 2002; Roberts, 1995), learners' noticing of feedback or interpretations of the targets of error corrections in recasts were examined; nevertheless, the coding systems used in the studies did not make it possible to investigate whether L2 learners notice the ungrammaticality of certain forms. Several researchers (Egi, 2007a; Sakai, 2004, 2011) attempted to devise coding systems by considering the types of linguistic data that L2 learners noticed in the recasts. Egi analyzed L2 learners' interpretation of feedback in terms of linguistic evidence for L2 learning. Her codes were (a) responses to content operationalized as "comments indicating that the learner interpreted the recast as the interlocutor's replies to the conversational content, such as confirmation and acknowledgement" (p. 522); (b) negative evidence, operationalized as "comments indicating that the learner was aware of error production or correction without a clear indication that the learner noticed the target-like form provided in the recast" (p. 524); (c) positive evidence operationalized as "comments indicating that the learner attended to the target-like model without reporting having produced errors or having been corrected" (p. 524); and, (d) negative + positive evidence operationalized as "comments that indicated the learner's (a) recognition of error production or correction and (b) attention to the target-like form in the recast" (p. 524). Her participants were 49 learners of Japanese. They performed two communicative tasks. Of the 49 participants, 31 participants, the immediate report group, were told to report their thoughts during the conversational interactions when they were prompted by the native-speaking interlocutor. The stimulated recall group consisted of 18 participants, who were told to report their thoughts after the communicative tasks. She found that the interaction data contained 2,017 feedback episodes, for 51.02% (1,029 episodes) of which the participants recalled their thoughts. Of the 1,029 episodes, 785 concerned recasts about morpho-syntactic items, and 134 concerned lexical items. The category *responses to content* accounted for 21.22% for the morpho-syntactic feedback episodes and 18.18% for the lexical feedback episodes; the category *negative evidence* accounted for 34.03% and 33.77%; the category *positive evidence* accounted for 18.70% and 18.80%; and the category *negative + positive* accounted for 26.05% and 29.87%. She further reported the effects of length of recasts and number of changes on the interpretation of recasts. Her findings showed that L2 learners interpreted longer recasts and recasts with more changes as responses to content more often than shorter recasts and recasts with fewer changes. Egi's detailed analysis of L2 learners' interpretations of recasts was of great significance in that she addressed the dual role of recasts as negative and positive evidence. Nevertheless, because the output activity (i.e., speaking or writing) itself can promote L2 learners' noticing (both noticing a hole and noticing the gap, Swain, 1993, 1995), Sakai (2011) pointed out that it is necessary to distinguish noticing errors through producing output from noticing errors through recasts. He pointed out that Egi's category system did not make a clear distinction between these two types of noticing. For example, her example for the category *negative + positive evidence* was shown as follows: "I had no idea how to say this verb, so I just say *hayai* ['fast'], which turned out to be wrong, and then you told me how to say 'hurry,' *isoide*" (p. 524). From this comment, this learner noticed his or her error at the moment of the production, and the recast should be interpreted as providing only positive evidence.

Sakai (2004, 2011) attempted to isolate the effect of production from that of recasts. Sakai (2004) examined the roles of output and feedback in L2 learners' noticing and repairs in later production. Sixteen Japanese learners of English (university students) were divided into two groups, the recast group and the model group. The recast group performed a picture description task in which the participants described pictures and received recasts on their errors. On the other hand, the model group performed two tasks (a picture description task and a picture identification task); in the first picture description task, the participants described pictures and received acknowledgments from the researcher; and in the second picture identification task, the participants identified the pictures as they listened to the researcher's descriptions, which were based on what they had produced in the first task and which served as models. Both groups performed the picture description task again. After the tasks, the stimulated recall session was carried out. Sakai devised the coding system for the stimulated-recall protocols and divided noticing into four categories: (a) problem, (b) production, (c) model/recast, and (d) a lack of noticing. The four categories refer to "noticing a hole in the interlanguage system" (p. 36), noticing "errors at the moment of production" (p. 36), "noticing the gap with the help of feedback [model or recast]" (p. 37), and unnoticing or noticing errors "at the moment of the interview" (p. 37). He found that both groups reported noticing errors at the moment of production (8.5% and 7.5%), which suggested that production helps noticing holes in the interlanguage. He also found that noticing the gap through feedback accounted for 3.9% (12 instances of the 307 errors for the model group) and 11.8% (30 instances of the 255 errors for the recast group) and that the recast group noticed more errors through recasts than the model group noticed errors through models. There were no statistically significant differences between the two groups in terms of the other three categories. The results suggest that recasts with the juxtaposition of the learners' erroneous utterances are more effective in getting learners to notice errors than models provided a few minutes after their utterances. Finally, the results show that noticing the gap through feedback (model and recast) was more effective in terms of repairs in later production than noticing a hole.

Utilizing the same coding categories for the stimulated-recall protocols as Sakai (2004), Sakai (2011) targeted irregular past tense forms in English and examined whether though recasts, L2 learners were provided with negative evidence that overgeneralized forms such as *drinked* and *shaked* are ungrammatical and positive evidence that irregular forms such as *drank* and *shook* are grammatical. His participants were 20 Japanese university students learning English, who were assigned to the recast group or the no-feedback group. They individually performed an information gap task with the researcher. The participants had a time line and the same set of 10 pictures that depicted Ken (a male character). They were told to place the pictures on the time line with the cards facing down. They were then told to turn over cards one by one and explain the cards to the researcher, who had the same timeline and picture cards. The researcher then had to place the cards in the same order as the participant. During this task, the participants in the recast group received recasts while the participants in the no-feedback group were provided with responses such as "Yes." "I see." and "OK." After the task, in the stimulated recall session, the participants listened to the recordings of the task interactions and were told to report their thoughts for every utterance. The results showed that the category *recast* for the recast group accounted for 21.6% of the 51 errors, which was significantly larger than that of the no-feedback group (no instances of that category); that the no-feedback group failed to notice statistically more errors than the recast group did; and that the two groups did not differ in the numbers of the categories *problem* and *production.* The results suggest that recasts effectively help L2 learners to recognize the ungrammaticality of overgeneralized forms.

In summary, Egi (2007a) clarified how L2 learners interpret recasts, in other words, whether they receive negative evidence, positive evidence, or negative and positive evidence. However, she did not differentiate noticing errors at the time of production (before the provision of recasts) and noticing errors through recasts. Sakai (2004, 2011) examined whether L2 learners notice the ungrammaticality of certain forms only with the help of recasts. However, Sakai (2004) targeted various grammatical errors, and thus it is not clear that his study examined either whether L2 learners learn ungrammaticality of overgeneralized forms or whether recasts help learners to narrow overgeneralized rules. Sakai (2011) selected irregular past tense forms; thus, his study is of significance in that he suggested that recasts can help L2 learners learn the ungrammaticality of the overgeneralized forms. However, it must be pointed out that his target structure might not concern the third type of incorrect hypothesis proposed by White (1989a). The errors in the irregular past tense *–ed* can be categorized as the second type of incorrect hypothesis discussed by White. That is to say, the incorrect hypothesis (in this case, attachment of *–ed*) and the correct hypothesis (in this case, irregular forms such as *drank* and *shook*) do not overlap. According to White, an incorrect hypothesis can be replaced by a correct one, only with "positive evidence of the existence of the alternate forms [that is, irregular forms]" (p. 143). Thus, still, it remains unclear whether recasts help L2 learners narrow overgeneralized grammatical rules.

Summary

In the above sections, I have reviewed and discussed empirical studies using grammaticality judgment tests and/or elicited imitation tests, studies reporting decreases in interlanguage forms, and studies using utilizing introspective methods. To sum up, no researchers have examined whether L2 learners can narrow overgeneralized rules with the help of recasts. Furthermore, none of the studies reviewed addressed the relevance of recasts to the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis. Although researchers have been interested in whether recasts provide negative evidence, that is, information about the ungrammaticality of particular forms (e.g., Nicholas, Lightbown, & Spada, 2001), and researchers have pointed out that negative evidence is necessary for L2 learners to retreat from overgeneralized rules (e.g., Long, 1996; White, 1989), there has been little research addressing this issue directly.

Methodological Issues

In this section, I review studies of the instruments used to measure the knowledge of ungrammaticality: grammaticality judgment tests, oral production, and elicited imitation. These instruments are discussed below.

Grammaticality Judgment Tests

L2 learners must decide "whether a sentence is well-formed or *deviant* [emphasis added]" (Ellis, 1991, p. 162) on grammaticality judgment tests; thus, as Gass (1994) stated, "Through the use of grammaticality judgments, we can establish whether learners' grammatical knowledge includes information about the impossibility of particular forms" (p. 306). White (2007) also made a similar statement that the grammaticality judgment task is the most commonly used technique to examine knowledge of L2 grammaticality and ungrammaticality (p. 44). Moreover, Munnich, Flynn, and Martohardjono (1994) went on to state that the grammaticality judgment task has been the only way to measure knowledge of ungrammaticality in the field of UG-based research on L1 and L2 acquisition (p. 229). These quotations indicate that grammaticality judgment tests have been used to assess knowledge of ungrammaticality as well as grammaticality.

The reliability and validity of grammaticality judgment tests have been investigated (e.g., Chaudron, 1983; Ellis, 1991; Gass, 1994; Sorace, 1996). Several researchers have argued that grammaticality judgment tests measure L2 learners' linguistic competence, that is, implicit knowledge (e.g., Arthur, 1980; Gass, 1983); on the other hand, other researchers (e.g., Bialystok, 1979; Ellis, 2005) have pointed out that grammaticality judgment tests might measure metalinguistic knowledge or explicit knowledge. In particular, time limits and the grammaticality of sentences have been found to influence the utilization of explicit knowledge. Ellis (2005) administered elicited imitation, a narration task, a timed grammaticality judgment test, an untimed grammaticality judgment test, and a metalinguistic knowledge instrument to 111 learners of English and performed a factor analysis. He found that the performances on untimed grammaticality judgment tests and metalinguistic knowledge constituted one factor (see also, Han & Ellis, 1998, for similar results). In terms of the time that L2 learners needed for judgment, Bialystok (1979) argued that judging ungrammatical items requires the utilization of explicit knowledge. Ellis (1991) showed L2 learners' think-aloud protocols that suggested that explicit knowledge was utilized for judging ungrammatical items (pp. 178-179).

In the field of L2 acquisition research, it has been proposed that L2 knowledge is divided into two types of knowledge about the target language, implicit and explicit knowledge, and that implicit knowledge is responsible for communicative language use (e.g., Ellis, 1994, 2004, 2005; Krashen, 1982; Paradis, 1994). For example, Ellis (2004, 2005) distinguished between explicit knowledge and implicit knowledge in terms of (a) awareness (intuitive awareness or conscious awareness of linguistic norms), (b) type of knowledge (procedural or declarative knowledge), (c) systematicity and certainty of L2 knowledge (variable but systematic knowledge or anomalous and inconsistent knowledge), (d) accessability of knowledge (automatic processing or controlled processing), (e) use of L2 knowledge (access to knowledge during fluent performance or access to knowledge during planning difficulty), (f) self-report (nonverbalizable or verbalizable), and (g) learnability. Furthermore, Ellis (2005) argued that no matter what theoretical orientations a researcher chooses, the L2 acquisition researcher needs to distinguish between learners' implicit and explicit knowledge (p. 168). For the present study, I followed Ellis's suggestions by utilizing multiple elicitation tests to measure implicit knowledge as well as explicit knowledge. I used untimed grammaticality judgment tests as a measure of explicit knowledge.

Oral Production

A possible means to measure implicit knowledge is communicative tasks that elicit free and spontaneous production from learners. In other words, the observation that learners do not produce ungrammatical structures or forms suggests that such rules allowing ungrammatical structures or forms might not be part of the learners' linguistic representation. However, as a measure of the knowledge of ungrammaticality, communicative tasks might not be effective, as they do not reliably elicit target structures (Ellis, 1991, p. 163; Erlam, 2006, p. 466).

Elicited Imitation

As an additional means to measure implicit knowledge of ungrammaticality, I used elicited imitation, the use of which, however, is not universally accepted. Bley-Vroman and Chaudron (1994) questioned the validity and reliability of elicited imitation tests as measures of L2 language competence (p. 246). On the other hand, Munnich, Flynn, and Martohardjono (1994) argued for the use of elicited imitation tests to measure learners' knowledge of ungrammaticality without detecting explicit knowledge. In a similar vein, Erlam (2006) argued that elicited imitation tests are a practical way to measure implicit knowledge and can be used as an alternative to communicative tasks (p. 467). Based on previous studies, she claimed that elicited imitation is reconstructive in nature, that is, elicited imitation tests require the learner to process the stimulus input for meaning and then to reconstruct the meaning based on the learner's interlanguage grammar rather than parroting the input verbatim (p. 469). To ensure the reconstructive nature of elicited imitation tests, she proposed that reconstructive elicited imitation tests of implicit grammatical knowledge have the following design features: (a) The test should require learners to pay attention to meaning rather than to form, (b) The test should set a time interval between the stimulus input and the response, and (c) Time pressure should be given so that learners cannot utilize explicit knowledge. Thus, Erlam (2006) claimed that elicited imitation can be a reliable and valid measure of implicit knowledge.

Summary

L2 learners' knowledge about ungrammaticality has been measured by means of grammaticality judgment tests, communicative tasks, and elicited imitation tests. As reviewed in this section, each type of measurement has advantages and limitations. Grammaticality judgment tests have been the most frequently used in the research on L1 and L2 acquisition. However, it has been pointed out that grammaticality judgment tests might tap explicit knowledge, not implicit knowledge. Thus, in order to test implicit knowledge, other measurements are required. One method to elicit L2 learners' performance reflecting implicit knowledge is free production elicited through communicative tasks. However, communicative tasks are not a reliable tool to elicit the target forms or structures. Another possibility is the use of elicited imitation tests. This type of measurement is more efficient than communicative tasks in eliciting the target forms or structures. However, not all the researchers agreed upon the reliability and validity of elicited imitation tests; some researchers (e.g., Erlam, 2006) argued that it is necessary to set test conditions to ensure the reliability and validity of the tests. Acknowledging each test's advantages and limitations, I followed Ellis's (2005) suggestions and used multiple elicitation tests (i.e., grammaticality judgment tests, oral production, and elicited imitation) to measure both implicit knowledge and explicit knowledge.

Definitional Issues

In this section, I discuss the definitions and variations of recasts in the literature. As Nicholas, Lightbown, and Spada (2001) pointed out (p. 749), researchers have defined and operationalized recasts differently. Defining recasts is a challenging undertaking.

Reformulation of Correct or Incorrect Utterances

Recasts have been defined as a reformulation of either correct or incorrect utterances. For example, Doughty (1994) defined a recast, a repetition, and expansion as a "response to an utterance that incorporates content words of the utterance, but also changes the utterance in some way (e.g., phonological, syntactic, lexical) but without adding any information" (p. 102), an "exact repetition of the learner utterance" (p. 102), and a "response to a learner utterance that provides additional information not contained in the learner utterance" (p. 102), respectively. Thus, according to her definition, recasts contain feedback concerning correct utterances as well as some toward erroneous utterances. Likewise, Long (1996) defined recasts as follows: "Recasts are utterances that rephrase a child's utterance by changing one or more sentence components (subject, verb, or object) while still referring to its central meanings" (p. 434). Again, his definition does not clearly state that a recast is a response only to an erroneous utterance by a learner. Furthermore, Long (1996, p. 434) pointed out that a recast entails four properties: (a) reformulation, (b) expansion, (c) semantic contingency, and (d) position (following the learner's utterance). In these four properties that a recast has, correction of errors does not play a role. Similarly, Bohannon, Padgett, Nelson, and Mark (1996), in a study of recasts for L1 learning, examined growth recasts, defined as "a broad set that includes all recasts that display relations between an initial child utterance (platform utterance) and a recast that expands, deletes, permutes, or otherwise changes the platform while maintaining significant overlap in meaning" (p. 551).

On the other hand, in most studies, recasts are considered reformulations of non-target-like utterances (Egi, 2010; Long, Inagaki, & Ortega, 1998; Long & Robinson, 1998; Mackey, 2012; Morris & Tarone, 2003). For example, Long, Inagaki, and Ortega (1998) used the term *corrective recasts* and defined them as

> ...responses which, although communicatively oriented and focused on meaning rather than form, incidentally reformulate all or part of a learner's utterance, thus providing relevant morpho-syntactic information that was obligatory but was

either missing or wrongly supplied, in the learner's rendition, while retaining its central meaning... (p. 358)

Similarly, Long and Robinson (1998) defined recasts as "corrective reformulations of a child's or adult learner's (L1 or L2) utterances that preserve the learner's intended meaning" (p. 23). Morris and Tarone (2003) defined a recast as "a listener's correct reformulation of the learner's incorrectly formed utterance" (p. 326). Egi (2010) defined recasts as "implicit negative feedback that reformulates learners' non-target-like utterances toward second language (L2) norms" (p. 1). Long (2007) defined a recast, what he called a corrective recast, as "a reformulation of all or part of a learner's immediately preceding utterance in which one or more non-target-like (e.g., lexical, grammatical) items is/are replaced by the corresponding target language form(s), and where, throughout the exchange, the focus of the interlocutors is on *meaning*, not language as object" (p. 77). Mackey (2012) defined recasts as "a form of feedback in which learners are provided with more target-like versions of their immediately preceding non-target-like utterances" (p. 14). These definitions clearly depict a recast as a correction of an ill-formed utterance.

For this study, I operationalized a recast as a reformulation of an incorrect utterance because this study focused on whether the feedback to deviant forms overgeneralized from the learners' L1 is effective in learning ungrammaticality of the overgeneralized forms. In other words, I was interested in the effectiveness of recasts given to incorrect forms that L2 learners make because of the L1 transfer.

Including L1 Use as Errors

Second, researchers differed in terms of what counts as errors; more specifically, whether errors involve the use of the L1 unsolicited by the interlocutor. For example, Lyster and Ranta (1997) considered recasts to include "translation in response to a student's use of the L1" (p. 47). Similarly, in the COLT (communicative orientation of language teaching) observation scheme, Spada and Fröhlich used the term *paraphrase* defined as "reformulation of previous utterance/s (including translation)" (p. 24). On the other hand, Doughty (1994) distinguished recasts from translation referred to as "immediate translation of learner utterance into French [L2]" (p. 102).

Whether L1 use is counted as errors is not relevant to this study. The target structures in this study were adjective-ordering rules and indirect passives. Thus, only syntactic errors were counted.

Addition of Emphasis

The third issue concerning the definition of recasts is the addition of emphasis. Although several researchers argued that recasts are implicit negative feedback (e.g., Long, 1996; Long & Robinson, 1998), other researchers (e.g., Ellis, 2007; Sheen, 2006) argued that, with particular features discussed below, recasts can be more or less explicit, that is, at some point along a continuum from implicit to explicit.

First, recasts can be accompanied by prosodic cues. For example, Nicholas, Lightbown, and Spada (2001) pointed out that recasts sometimes include stress or emphasis as well as a reformulation of a learner's erroneous utterance (p. 749).

Chaudron (1977) listed the type of feedback *REPETITION with CHANGE and EMPHASIS* defined as “T [teacher] adds EMPH [emphasis] to stress location of ERROR and its correct formulation” (p. 39). He gave an example as follows:

> S: Doo tout …
> T: D*u* tout. (stress)

This category can be considered a recast with prosodic emphasis.

Emphasis takes other forms also. For example, in Doughty and Varela (1998), recasts were made salient, preceded always by repetition of an erroneous utterance. They called this kind of recasts *corrective recasts*.

For this study, recasts were provided naturally, that is to say, without addition of emphasis. This is because this study aimed at testing Leeman’s (2003) enhanced salience hypothesis. As reviewed in the Enhanced Salience Hypothesis section, she claimed that the juxtaposition of recasts with erroneous utterances helps increase the salience of the input and that the enhanced salience is facilitative of L2 learning. In order to test whether the enhanced salience of the recasts due to their juxtaposition with erroneous utterances are effective for L2 learners to learn the ungrammaticality of certain orders that L2 learners consider to be acceptable because of the L1 influence, recasts in this study were provided with no addition of emphasis.

Addition of Information

Fourth, recasts differ in terms of adding information to the original utterance or reducing the original information. For example, Lyster (1998a) used the isolated/incorporated distinction for his categories of recasts. An isolated recast refers to the reformulation of all or part of the learner's utterance (p. 58); on the other hand, an incorporated recast refers to the incorporation of all or part of the learner's utterance into a longer statement (pp. 58-59). Roberts (1995) made a distinction between recasts and partial recasts. The latter was defined as "Similar to the preceding [recast], but teacher only models the segment of the utterance in which the error occurs" (p. 170). Chaudron (1977) posited the categories *REDUCTION* and *EXPANSION* as well as *REPETITION with CHANGE.* Similarly, Sheen (2006) identified three types of repetitions of the original utterance: (a) partial/reduced recasts, (b) complete/full recasts, and (c) expanded recasts (p. 366). Thus, a recast can be part or all of the original utterance with changes and can be an utterance that incorporates the reformulation of part or all of the original utterance.

The addition or reduction of information might influence the explicitness of the recast. For example, Ellis, Loewen, and Erlam (2006) argued that partial recasts can help L2 learners' locate the errors (p. 342). Sheen (2006) found a statistically significant difference in learners' repair between reduced recasts and full recasts and suggested that reduction increased learners' noticing of errors (p. 384).

Some previous researchers (e.g., Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Leeman, 2003) utilized partial recasts whereas other researchers (e.g., Doughty & Varela, 1998; Loewen & Nabei, 2007; Sakai, 2004, 2007) adopted full recasts. A choice of recast type might depend on the target structures. Studies focusing on morphological rules (the regular past tense *–ed*, the comparative *–er*, and noun-adjective agreement in gender and number) used partial recasts; on the other hand, studies focusing on syntactic rules (the past conditional and question formation) utilized full recasts. The choice of full recasts in Sakai (2004, 2007) was derived from their research design in which recasts were compared to non-corrective repetitions.

For this study, a recast was operationalized as a reformulation of all of an incorrect utterance; in other words, full recasts were provided. One main reason for this was that this study focused on syntactic rules (adjective-ordering rules and indirect passives). In particular, it was difficult to segment indirect passives and give relevant partial recasts. Another reason was that this study attempted to control the amount of input across the conditions.

Intonation

Fifth, recasts can be provided with rising intonation or with falling intonation. With the isolated/incorporated distinction, Lyster (1998a) categorized recasts into four types as follows: (a) isolated declarative recast, (b) isolated interrogative recast, (c) incorporated declarative recast, and (d) incorporated interrogative recast.

When recasts are accompanied by rising intonation, they are also referred to as confirmation checks. Long (1980) defined a confirmation check as “any expression by the NS [native speaker] immediately following an utterance by the interlocutor which was designed to elicit confirmation that the utterance had been correctly understood or correctly heard by the speaker” (p. 81). He also stated “Confirmation checks are always formed by rising intonation questions, with or without a tag” (pp. 81-82).

As Oliver (1995, 2000) pointed out, recasts and confirmation checks can overlap. Oliver (2000, p. 121) stated that the feedback in the following example was considered both a recast that contains two changes (*have* and *cups*) and a confirmation check.

NNS:	And I'ave two-two cup.	
NS:	You have two cups?	[a recast and a confirmation check]
NNS:	Yeah.	

On the other hand, the following feedback was a recast, not a confirmation check.

NNS:	I think a Perth.
NS:	You think it's Perth.

As the above examples indicate, confirmation checks and recasts overlap when they accompany rising intonation and follow a learner's erroneous utterance. Recasts with falling intonation are not considered confirmation checks. Further, confirmation checks do not always contain a reformulation of erroneous utterances.

For this study, recasts were always provided with rising intonation. This is mainly because recasts with rising intonation take the conversational role of confirmation checks and show the recast provider's intention for mutual understanding.

Length of Recasts and the Number of Changes

Sixth, several researchers (e.g., Doughty, 1994; Egi, 2007a; Philp, 2003) have divided recasts in terms of (a) the length of recast, that is, shorter recasts (five morphemes or less for English and seven or less for Japanese) and longer recasts (six morphemes or more for English and eight or more for Japanese), and (b) the number of changes in the recasts. Previous studies reported that these variations affected noticing of errors or uptakes in the learners' turn following the feedback (e.g., Doughty, 1994; Egi, 2007a; Philp, 2003).

For this study, the length of the recasts and the number of changes were not controlled because this study targeted at two different structures. Recasts for adjective ordering were shorter than those of indirect passives; recasts for adjective ordering tended to involve a fewer number of changes than those of indirect passives. Thus, these characteristics were not controlled for this study.

Frequencies of Recasts

Seventh, how frequently recasts are provided in a certain period of time can differ. The frequencies of recasts might not be related directly to the definition of a recast; rather it concerns the condition of the experimental treatments. For example, the recast groups in the study of Mackey and Philp (1998) received intensive recasts of non-target-like utterances; on the other hand, Lyster (1998a, 1998b) investigated naturally occurring recasts in classroom interactions.

The difference in the frequencies of providing recasts might reflect the differences in the research settings varying from more controlled experiments to more natural classrooms. Mackey (2012) pointed out that "in tightly controlled laboratory contexts where linguistic form is clearly the focus, learners are able to attend to the feedback that is provided and incorporate it into their developing interlanguages" (p. 48).

For this study, recasts were provided frequently in a short period of time. This is because this study was an experimental study in nature and controlled the provision of recasts. Recasts or non-corrective repetitions were provided to all the utterances of L2 learners. Thus, the frequencies of recasts were assumed to be relatively high.

Summary

In sum, different researchers have defined recasts differently. Thus, as Nicholas, Lightbown, and Spada (2001) pointed out (p. 749), different definitions of recasts make it difficult to make comparisons across studies although a large number of researchers have investigated recasts. Several researchers compared different types of recasts and found the differential effectiveness of various types of recasts (e.g., Doughty, 1994; Egi, 2007a; Sheen, 2006; Philp, 2003). Mackey and Goo (2007) acknowledged this point; however, in their meta-analysis, they treated recasts as one category for practical reasons (p. 413). At the moment, for research on recasts, it is important to operationalize recasts and provide detailed information about recasts in terms of their corrective nature, stress and emphasis, segmentation/repetition/expansion, intonation, the length of recasts and the number of changes, and the frequencies.

Gaps in the Literature

Two gaps exist in the literature. First, as far as I am aware, no researchers have examined the effects of recasts on the retreat from overgeneralized grammar, that is, whether recasts are effective for the acquisition of knowledge that overly general rules are ungrammatical. Several studies (e.g., Ellis, 2007; Ellis, Loewen, & Erlam, 2006) utilized the grammaticality judgment tests; however, they did not investigate whether their learners obtained knowledge about the ungrammaticality of their target forms, but rather examined whether they improved in the control of the application of the knowledge about the ungrammaticality of the structures. Further, as I reviewed in the Empirical Studies section, no previous studies have shed light on learnability problems. None of the studies have focused on the structures that constitute superset-subset relationships (or at least overlapping relationships) between the learner language and the L2 and that are assumed to be difficult to acquire only with positive evidence.

Second, and more importantly, to my knowledge, so far no previous researchers have tested the three theoretical hypotheses (the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis) in the same study. These hypotheses provide different explanations of the cognitive mechanisms underlying recasts. However, the question of which hypothesis is empirically supported remains unaddressed.

Purposes of the Study

This study experimentally isolated recasts (juxtaposed with learner errors), non-contingent positive evidence (not juxtaposed with learner errors), recasts plus additional input, and enhanced salience of positive evidence as independent variables. Corresponding to each of these variables, four treatment groups were created: the Recast Group (the REC Group), the Non-Contingent Positive Evidence Group (the POS Group), the Recast Plus Additional Input Group (the REC+ Group), and the Input With Enhanced Salience Group (the SAL Group). In this study, I investigated the effects of four types of interactional input on the L2 learning of English adjective-ordering rules and indirect, or adversative, passives.

There are two main purposes for the current study. The first purpose was to examine the effects of recasts on obtaining the knowledge that overgeneralized rules from L1 (i.e., Japanese) are ungrammatical in English (adjective-ordering rules and indirect passives). Several researchers (e.g., Gass, 1997; Long, 1996; White, 1987, 1989a) have argued that L2 learners need to receive negative evidence in order to retreat from overgeneralized rules that are formed mainly because of L1 influences and that cannot be eliminated only with positive evidence. However, so far little light has been shed on the acquisition of information about what is not possible in the L2. Thus, I attempted to examine (a) whether L2 learners created overgeneralized rules from Japanese about adjective-ordering rules and indirect passives and would not retreat from overgeneralized rules only with the non-contingent positive evidence and (b) whether recasts would help L2 learners learning that overgeneralized rules are ungrammatical.

The second purpose was to test three hypotheses, the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis. This purpose is important because testing the three hypotheses makes it possible to examine their different explanations about how recasts facilitate L2 learning. Thus, I attempted to compare the effects of recasts with (a) recasts plus additional input and (b) positive evidence with enhanced salience.

Research Questions and Research Hypotheses

The general research questions are: (a) Are recasts effective for narrowing overgeneralized rules in an L2, and (b) if so, how do recasts work? Based on the theoretical hypotheses reviewed above, five specific research questions and hypotheses are posited for this study: In relation to learning the ungrammaticality of overgeneralized rules,

1. Is exposure to non-contingent positive evidence alone effective? It is hypothesized that the POS Group will not gain from the pretest to the posttests.
2. Is exposure to input with recasts (juxtaposed with learner utterances) more effective than exposure to non-contingent positive evidence (not juxtaposed with learner utterances)? It is hypothesized that the REC Group will perform better than the POS Group.
3. Is exposure to input with both recasts and additional subsequent input more effective than exposure to input with recasts? It is hypothesized that the REC+ Group will perform better than the REC Group.
4. Is exposure to input with enhanced salience of positive evidence (without negative evidence) more effective than exposure to input with unenhanced positive evidence (non-contingent positive evidence)? It is hypothesized that the SAL Group will perform better than the POS Group.
5. Does exposure to input with recasts differ from exposure to input with enhanced salience of positive evidence? It is hypothesized that no differences will be found between the REC and SAL Groups.

The first hypothesis is predicted by White's (1989a) argument that positive evidence alone does not lead to retreat from overgeneralized rules. Disconfirmation of this hypothesis suggests that positive evidence would have functioned as indirect negative evidence. Indirect negative evidence is the idea that learners might learn the ungrammaticality of certain forms or structures on the basis of the observation of the non-occurrence of the forms or structures. However, White argued, "This proposal [the idea of indirect negative evidence] runs into difficulties because of its vagueness" (p. 15). Furthermore, empirical studies (Izumi & Lakshmanan, 1998; Trahey, 1996; Trahey & White, 1993; White, 1989b, 1991) showed that positive evidence alone does not lead to learning of the ungrammaticality of overgeneralized forms and that negative evidence is needed. Thus, the first hypothesis is that the POS Group will not gain from the pretest to the posttests. The interpretations of the results for the hypotheses two to five will be based on the assumption of non-gain of the POS Group.

The direct contrast hypothesis (Saxton, 1997, 2000) predicts the second hypothesis. The REC Group will receive recasts that are contingent to and juxtaposed with the learners' errors; on the other hand, the POS Group will receive positive evidence about the target structures that is not juxtaposed with the learners' errors. According to the direct contrast hypothesis, the juxtaposition will help the learners perceive negative evidence; thus, the direct contrast hypothesis predicts better performance by the REC Group than the POS Group.

The third hypothesis is generated from the additional input hypothesis (Gass, 1997; Gass & Mackey, 2007; Long, 1996, 2007). The additional input hypothesis states that additional input is facilitative for L2 learners to modify or reformulate their interlanguage hypotheses; thus, it predicts that the REC+ Group will perform better than the REC Group in relation to learning the ungrammaticality of overgeneralized rules.

The interpretations of the results regarding the second and third research hypotheses will enable me to test the direct contrast and the additional input hypotheses. First, if the second and third research hypotheses are confirmed, then both the direct contrast hypothesis and the additional input hypothesis will be supported. In other words, the findings will be interpreted as suggesting that the juxtaposition is a necessary and sufficient condition for recasts to work as negative evidence and that recasts are more effective with additional input. Second, if the second research hypothesis is not confirmed but the third one is supported, then the findings will support the additional input hypothesis only. In other words, juxtaposition is not regarded as a sufficient condition, and the additional input provided after the provision of recasts is necessary. Lastly, if both research hypotheses are rejected, neither the direct contrast hypothesis nor the additional input hypothesis will be supported.

The fourth and fifth research hypotheses concern the enhanced salience hypothesis (Leeman, 2003), which predicts that recasts and positive evidence with enhanced salience will work in the same way. Thus, it is predicted that the SAL Group will perform better than the POS Group in relation to learning the ungrammaticality of the overgeneralized rules and at the same time that the SAL Group will perform as well as the REC Group. If these hypotheses are disconfirmed, Leeman's enhanced salience hypothesis will have to be interpreted to be limited to learning some structures or forms, not to concern retreat from overgeneralized rules. If the hypotheses are supported, the enhanced salience hypothesis can be considered to apply to learning the ungrammaticality of certain structures or forms as well.

CHAPTER 3

METHODS

Participants

The participants in this study were 97 Japanese university students learning English. They volunteered to take part in this study. They were recruited in several ways. First, I explained the research to students in my classes (Academic English II, First-Year Seminar for English Education Majors, An Introduction to Teaching English, Teaching of English Oral Communication Skills I, Teaching of English Oral Communication Skills II, Theories for Teaching as a Foreign Language I, Theories for Teaching as a Foreign Language II, and Teaching English to Young Learners) and in other classes (Educational Computing) in order to ask them to participate in the study. Of the 97 participants, 70 university students were recruited from the classes at one university. Second, I asked those who had taken my classes to take part in this study (n = 8). Third, 14 participants were introduced to me by five participants and volunteered to participant in this study. Finally, two acquaintances of mine, who taught at a university and a national college of technology respectively, introduced their students to me (n = 5). Therefore, two types of relationship with me were observed: 75 participants the teacher-student relationship with me whereas 22 had the researcher-participant relationship with me.

All of the participants were asked to fill in the written consent form (see Appendix A). After completing all the experimental sessions, they were paid 2,000 yen.

The participants were randomly assigned to four treatment groups: the Recast Group (REC Group), the Non-Contingent Positive Evidence Group (POS Group), the Recast Plus Additional Input Group (REC+ Group), and the Input With Enhanced Salience Group (SAL Group). The group assignment was carried out as follows: First, participants were ordered using random digits, and then they were assigned to groups in turn.

According to the background questionnaire (see Appendix B), two participants lived abroad for more than one year. One stayed in the United States for two and a half years before the age of five. Although he attended a local kindergarten, he mentioned that his main language at home was Japanese. The other participant lived in Peru for three years from Grade 1 to Grade 3. She went to a Japanese school in Peru. Because these two participants reported that their main language was Japanese and that they studied English at school in Japan, their data were included in the statistical analyses in this study.

The 97 participants were screened for each target structure (adjective ordering and indirect passives) in several steps. The screening for adjective ordering was as follows. First, of the 97 participants, 17 in the REC and the REC+ Groups did not receive any recasts about adjective ordering in the treatment sessions. Because the provision of recasts depend on the participants' production of errors, no provision of recasts about the target structures indicate that they did not make errors about the target structures. They were excluded because it was evidence that the participants had already learned the structure. Second, in the follow-up questionnaire (Appendix C), four reported at the end of the experimental session that they had had a chance to study the target structures during the period between the pretest and the delayed posttest by themselves or in linguistics classes. These four participants' data were not analyzed. Lastly, the performance of one participant on one test was found to be unrecorded because of a problem with the recorder. These 22 participants were excluded from the adjective-ordering analysis.

The exclusion procedure for indirect passives was as follows. Of the 97 participants, six in the REC and the REC+ Groups did not receive any recasts about indirect passives in the treatment sessions and were not included in the analysis because it was evident that they participants had already learned the structure. In the follow-up questionnaire (Appendix C), one participant reported at the end of the experimental session that she had had studied the target structures during the period between the pretest and the delayed posttest. Her data were excluded from the analysis. Thus, these seven participants were excluded from the indirect passives analysis.

Table 1 shows the final number of the participants for each group. The number of participants for adjective ordering was 75 (68 whose data were analyzed for both structures and 7 whose data were analyzed only for adjective ordering): the POS Group (n = 20), the REC Group (n = 20), the REC+ Group (n = 16), and the SAL Group (n = 19). The number of participants for indirect passives was 90 (68 whose data were analyzed for both structures and 22 whose data were analyzed only for indirect passives): the POS Group (n = 21), the REC Group (n = 22), the REC+ Group (n = 27), and the SAL Group (n = 20).

Table 1. *The Number of Participants in Each Group*

Group	POS	REC	REC+	SAL	Total
Analysis					
ADJ & IDP	20	16	13	19	68
ADJ only	0	4	3	0	7
IDP only	1	6	14	1	22
Total	21	26	30	20	97

Note. ADJ & IDP = those whose data were analyzed for adjective ordering and indirect passives; ADJ only = those whose data were analyzed only for adjective ordering; IDP only = those whose data were analyzed only for indirect passives. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

According to the background questionnaire (Appendix B), all participants were L1 speakers of Japanese. Other information about the participants was summarized in Table 2. Of the 97 participants, 54 were female, and 43 were male. Most of the participants were studying in the undergraduate program of a College of Education at a national university (University S) in central Japan (*n* = 92), the hensachi of which is 57; one was an undergraduate student in the faculty of letters of a private university (University H) in Tokyo, the hensachi of which is 62; and four were fourth-year students at a national college of technology (College N) in central Japan. National colleges of technology are one school type that provides five-year engineering education after junior high schools. The fourth-year at a national college of technology is equivalent to first-year at a university. They were first-year to fourth-year students (eight first-year students, 44 second-year students, 18 third-year students, and 27 fourth-year students). They were 20 years and 9 months old on average. All of them reported in the questionnaire that they had learned English mainly through English education in Japanese junior and senior high schools for six years. The length of English language study was 9 years and 2 months on average. This length was considered to include English language study at elementary schools and at universities as well as the six-year study at junior and senior high schools. In Japan, teaching foreign language conversion (mainly, English language) started as a part of international understanding during the Period for Integrated Study for Grades 3 to 6 at the elementary schools in 1998 and aimed at providing experiential learning activities in which elementary students were exposed to English and could be familiarized with foreign cultures (Ministry of

Table 2. *Participants' Background Information*

	Group				
	POS	REC	REC+	SAL	Total
n	21	26	30	20	97
Gender					
Female	11	12	17	14	54
Male	10	14	13	6	43
School					
University S	20	25	29	19	92
University H	0	0	0	1	1
College N	1	1	1	1	4
Grade					
1st year	1	2	4	1	8
2nd year	10	12	13	9	44
3rd year	3	6	5	4	18
4th year	7	6	8	6	27
Age					
M (SD)	20;9 (1;2)	20;6 (1;1)	20;9 (2;3)	21;3 (2;0)	20;9 (1;9)
English Study					
M (SD)	9;1 (2;8)	8;9 (1;10)	9;3 (2;4)	9;5 (1;11)	9;2 (2;2)

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Education, Cultures, Sports, Science and Technology, 2001, pp. 121-125). Therefore, the participants were considered to have started their English language study in elementary school.

In order to examine whether the participants in the four groups differed in their grammatical knowledge, the structure section of the Comprehensive English Language Test for Learners of English (CELT Form A, Harris & Palmer, 1986) was administered. The test consisted of 75 multiple-choice questions and took the participants approximately 45 minutes to complete. Table 3 provides the results of the structure section. As to the grouping for adjective ordering, the reliability of the test was .89. A one-way ANOVA performed on the results showed no statistically significant difference among the groups: $F(3, 69) = 0.16$, $p = .92$. As to the grouping for indirect passives, the reliability was .87. Another one-way ANOVA was performed and showed no statistically significant difference among the group, either: $F(3, 84) = 0.55$, $p = .65$; thus, the participants in the four groups had similar grammatical knowledge.

Table 3. *CELT (Structure) Results by Group*

	Total		ADJ		IDP	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Group						
POS	19	41.63 (12.69)	18	40.83 (12.55)	19	41.63 (12.69)
REC	26	41.62 (11.06)	20	41.55 (12.18)	22	39.45 (10.07)
REC+	30	43.00 (9.95)	16	42.56 (11.24)	27	41.74 (9.18)
SAL	20	43.75 (11.89)	19	43.42 (12.12)	20	43.75 (11.88)

Note. CELT scores were not obtained from four participants of the POS group. ADJ = those whose data were analyzed for adjective ordering; IDP = those whose data were analyzed for indirect passives. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

On the basis of the information obtained from the background questionnaire (Appendix B), the meta-language test (Appendix D), the exit questionnaire (Appendix E), and the follow-up questionnaire (Appendix C), the participants' information by group for the two target structures was summarized in Appendices F and G. The four groups did not differ in terms of the relationship with the researcher (REL), the number of those who had experience abroad (EXP), previous knowledge about the ungrammaticality of the color-size-noun sequence and indirect passives in English (PREV), the scores of the meta-language tests that were administered after the delayed posttests (MLT), and participant's awareness about the purpose of the study (EXIT). Fisher's exact tests did not show statistically significant differences among the four groups for the first four points; one-way ANOVAs did not show that the four groups differed significantly in the participant's awareness about the purpose of the study.

Research Design

A pretest, posttest, and delayed-posttest design was employed. The independent variable was the treatment conditions (four groups); the dependent variable was the test scores regarding ungrammaticality, measured by three testing tasks (oral production, elicited imitation, and grammaticality judgment).

Table 4 shows the experimental schedule of this study. In Session 1, the participants were provided with an explanation about the study and asked to sign the consent form (Appendix A). They then answered the background questionnaire (Appendix B) and took the CELT test (structure section), and the pretest. About one week later, they took part in Session 2 in which they were provided the treatment and took the immediate posttest. Approximately one month later, in Session 3, they took the delayed posttest. The participants also took the meta-language test in which they had to correct three errors relating to two target structures (adjective ordering and indirect passive) and one distractor (irregular past form) and to explain the corrections as well as to make judgments (see Appendix D). They also completed the exit questionnaire (Appendix E). In addition, the participants were asked to fill in the follow-up questionnaire (Appendix C) after the completion of the three sessions. Five of the 97 participants did not answer the follow-up questionnaire (one for the POS Group, one for the REC Group, two for the REC+ Group, and one for the SAL Group).

Table 4. *Experimental Schedule*

Session 1	➔	Session 2	➔	Session 3
Explanation about the study Consent form Background questionnaire CELT test (45 min.) Pretest (30 min.)	One-week interval	Treatment tasks Immediate posttest	One-month interval	Delayed posttest Meta-language test Exit questionnaire

The aim of the provision of delayed posttest was to test the durability of any learning effects and to detect the delayed effects of the treatment. This is because several researchers have reported delayed effects of interaction on L2 development (e.g., Gass, 1997; Mackey, 1999, 2012; Mackey & Goo, 2007).

Target Structures

Two structures were chosen for the present study. One reason for that choice is that Ellis (2007) called for studies targeting different structures (p. 360), arguing that the effects of feedback differ for different structures. Another reason is that findings about two different structures increase their generalizability. The main criteria for selecting the target structures were as follows: (a) Japanese learners of English face a learnability problem regarding the structures, (b) they make persistent errors when producing or comprehending the structures, and (c) the structures are within the L2 learners' processing capabilities. Adjective-ordering rules and indirect passives were selected for this study. In the following sections, each structure is explained in details, and the reasons for the choice of the structures are provided according the above criteria.

Adjective-Ordering Rules

Adjective-ordering rules are governed by semantic constraints in English. Among previous studies on L2 acquisition, Long, Inagaki, and Ortega (1998) selected adjective-ordering rules as the target structure. In English, the order of premodifying descriptive adjectives is normally fixed: epithet (i.e., characteristic)–size–shape–age–color–origin–substance (i.e., material) (Close, 1975, p. 159; Declerck, 1991, p. 350; Swan, 1995, p. 8). As Swan (1995) pointed out, "different grammars disagree about the details [of adjective order]" (p. 8). For example, Thomson and Martinet (1986) showed a different order of adjectives: size–general description–age–shape–color–material–origin–purpose (p. 35). Although several variations exist among researchers, disagreement does not exist for the order of size and color. English normally allows size-color-noun sequence only, like *a large green Chinese carpet* (Close, 1975, p. 159).

Celce-Murcia and Larsen-Freeman (1999) pointed out that L2 learners of English might produce errors in the order of more than two adjectives because “not all languages follow a prenominal order the way English does” (p. 392). They showed that Arabic learners of English produce errors such as *an American interesting movie* and *a wooden big bowl*. Similar to Arabic, the Japanese language has a less fixed adjective order. Yasui (1996) pointed out this difference between Japanese and English and stated that the Japanese language allows *akai–ookina–hon* (red-large-book) and *ookina–akai–hon* (large-red-book), while only the latter sequence, *a large red book,* is permitted in English (p. 121). In other words, the Japanese language allows both a color-size-noun sequence and size-color-noun sequence (e.g., *akakute–chiisai–inu* and *chiisakute–akai–inu* meaning “a small red dog”).

According to the Courses of Study (the official guideline of the curricula) issued by Ministry of Education, Culture, Sports, Science and Technology (2008, 2010), adjectives are introduced at junior high schools. However, adjective ordering rules are not stipulated as target structures for junior and senior high school students. Thus, it is not clear whether Japanese learners of English learn adjective ordering rules.

The color adjectives used in the tests and the treatments were *black*, *blue*, *brown*, *gray*, *green*, *orange*, *pink*, *red*, *white*, and *yellow* (k = 10). The adjectives of size used in the tests and the treatments were *big*, *high*, *large*, *little*, *long*, *short*, *small*, and *tall* (k = 8). The color adjectives were ordered alphabetically and numbered. The adjectives assigned odd numbers (*black*, *brown*, *green*, *pink*, and *white*) were used for the treatment, whereas those assigned even numbers (*blue*, *gray*, *orange*, *red,* and *yellow*) were used only for testing. The size adjectives were divided into old items (*big*, *little*, *short*, and *tall*) used for both testing and the treatment and new items (*large*, *small*, *long,* and *high*) used only for testing. Every test sentence consisted of new combinations of items that were designed to test whether the participants generalize what they had learned to new combinations of items. The nouns used for the testing and treatment were high frequency words that were selected from junior high school textbooks. The test sentences were run though Vocabprofiler 2.7 (Cobb, 2006) to check their frequency levels; all of the words were in the high frequency 2,000 words of English except for *computer*, which was in the Academic Word List (Coxhead, 2000).

Reasons for Selecting Adjective Ordering

The first reason that this structure was chosen is because it poses a learnability problem for Japanese learners of English. It is predicted that Japanese learners of English will transfer L1 rules to the L2 and judge both adjective orders as acceptable in English. By receiving positive evidence, that is, models of the size–color–noun sequence in English, they can recognize that size–color–noun is permitted in English and might increase their use of that sequence. However, this does not mean that they will learn that color–size–noun is not acceptable in English. Therefore, it is hypothesized that negative evidence is necessary to eliminate the rules overgeneralized from the Japanese language.

Second, Japanese learners of English tend to make persistent errors in the structures and are likely to accept the color–size–noun sequence in English (e.g., Sakai, 2003). Sakai conducted three experiments at Japanese junior high schools to examine whether Japanese learners of English transfer their L1 to the L2 system and accept the color–size–noun sequence, whether positive evidence only helps L2 learners learn ungrammaticality of that sequence, and whether focus on meaning and focus on form lead to differential effects. In Experiment 1, 49 junior high school students in Grades 2 and 3 (equivalent to Grades 8 and 9) were divided into three groups: the Focus on Form (FonF) group (n = 16), the Focus on Meaning (FonM) group (n = 16), and a control group (n = 17). On the first day, the three groups performed the grammaticality judgment test that consisted of five grammatical sentences and five ungrammatical sentences. Thus, the maximum score was 5 for grammatical and ungrammatical sentences respectively. The treatment lasted for three days (from the second day to the fourth day). On each day, the participants in the FonF group listened to 10 sentences that include the color–size–noun sequence two times (e.g., *I have a small white car*). On the first listening, the participants were required to choose the appropriate translations of the sentences among the three written choices (e.g., *kiiroi-kuruma* [a yellow car], *ookii-kuruma* [a big car], and *shiroi-kuruma* [a white car]); and on the second listening, they were told to select one of the two English sentences (e.g., *I have a big yellow car* and *I have a small white car*). Thus, the task was designed so that the participants could attend the target part (e.g., color–size) of the sentence while they focused on the meaning. The participants in the FonM group listened to the same sentences two times.

The first listening task and the second listening task were the same for the FonM group so that the participants could attend the messages of the sentences on the both listening times. The control group was not provided with the listening tasks. On the fifth day, the three groups took the grammaticality judgment test. Experiment 2 was carried out in one class hour (i.e., 50 minutes). Eighty-two junior high school students in Grades 2 and 3 were allotted to one of the three conditions: the explicit instruction group (n = 29), the intensive FonF group (n = 26), and a control group (n = 27). The explicit instruction group was provided with 15-minute explicit explanation about the differences between the Japanese and English adjective ordering rules. The intensive FonF group was provided with the listening task in which the participants listened to 30 sentences two times under the same conditions as the FonF group of experiment 1. In Experiment 3, 59 junior high school students in Grades 2 and 3 were divided into two groups: the FonF group (n = 34) and the FonM (n = 25). The procedures were the same as Experiment 1. In general, the results showed that on the pretest, the learners' performance on grammatical sentences was relatively high (e.g., M = 4.26 for the FonF group in Experiment 3; M = 3.84 for the FonM group in Experiment 3) whereas their performance on the ungrammatical sentences was quite low (e.g., M = 1.21 for the FonF group in Experiment 3; M = 1.36 for the FonM group in Experiment 3, suggesting that Japanese learners had a tendency to accept both size–color–noun and color–size–noun sequences, which could be attributed to L1 transfer. In addition, the results of Experiments 1 and 3 showed that even after a three-day intensive provision of the target structure (10 sentences per day), the learners' performance on the grammaticality judgment tests (for both grammatical and ungrammatical sentences for Experiment 1 and for ungrammatical sentences for Experiment 3) did not improve. For example, in Experiment 1, the FonM did not improve from the pretest (M = 3.63 for grammatical sentences; M = 1.94 for ungrammatical sentences) to the posttest (M = 3.69 for grammatical sentences; M = 1.06 for ungrammatical sentences). Furthermore, the results of Experiment 2 showed that an intensive exposure to the target structure (30 sentences at one time) did not lead to improvement on the grammaticality judgment test (from 4.08 to 4.00 for grammatical sentences and from 1.23 to 1.54 for ungrammatical sentences). The results of Experiment 2 showed that participants in the explicit instruction group improved on the ungrammatical sentences (from 1.10 to 3.62 on average) and that they rejected the ungrammatical color–size–noun sequence successfully on the posttest. Thus, these results suggest that positive evidence does not work for learning of ungrammaticality of the color–size–noun sequence in English and that L2 learners might require negative evidence to retreat from their overgeneralized hypothesis about the adjective ordering.

Because the participants in Sakai (2003) were young learners of English, I checked whether older learners of English make persistent errors in the structures and are likely to accept the color–size–noun sequence in English as well. I examined the results of one question item in the exit test administered after the two-year required English course. The test was administered to second-year university students in the Faculty of Education in University S in January, 2008. The test included the following multiple-choice question.

[Question] “What did you buy yesterday?” “I bought ( ).”

[Choices] 1. two black small bags 2. black small bags two

3. two small black bags 4. small black bags two

The total number of students was about 260. Because 28 students did not agree to participate on the consent form, the responses from 232 students were analyzed. Table 5 presents the results. Of the 232 students, 24.1% chose the sequence of color-size-noun, whereas 70.3% chose the sequence of size-color-noun. The results suggest that some Japanese university students overgeneralized the Japanese adjective-ordering rules to English and accept the color-size-noun sequence.

Table 5. *Frequency and Rate of Students’ Choice of Each Alternative (Adjective Ordering)*

	Frequency	Rate
1	56	24.1%
2	3	1.3%
3	163	70.3%
4	7	3.0%
No response	3	1.3%

Third, the rules of adjective ordering are processable within the learner interlanguage system (Pienemann, 1998a, 1998b, 2003). Even if L2 learners notice the mismatch of their utterances and the target language, they might have difficulty in incorporating and processing the forms and structures beyond their developmental stages (e.g., Sakai, 2004). Thus, the difficulty of the target rules should correspond to L2 learners’ proficiency level (e.g., Long, Inagaki, & Ortega, 1998). Processability theory (Pienemann, 1998a, 1998b, 2003) proposes five language-specific processing procedures: (a) word/lemma access, (b) the category procedure, (c) the phrasal procedure, (d) the S-procedure, and (e) the subordinate clause procedure. Language learners are hypothesized to acquire these procedures in turn in this order. Adjective ordering requires the use of the second procedure, phrasal, which is considered a relatively low-level procedure.

Indirect Passives

The second target structure was the indirect passive. The indirect passive has been targeted in previous research on Japanese-speaking learners of English (Izumi & Lakshmanan, 1998; Suzuki, 1996) and English- and Chinese-speaking learners of Japanese (Hara & Ma, 2004).

According to Abe (1999), the passive sentence has the following characteristics: The sentence (a) involves a transitive verb, (b) places a direct object (i.e., the internal argument of the verb) in the subject position of the sentence, (c) adds a morphological change to the verb, and (d) optionally expresses an agent of the verb as an adjunct. Abe explained that in the government and binding theory version of generative grammar, the addition of the morphological change to the verb, for example, *-en* forms in English, leads to the suppression of the theta role of the external argument and the absorption of the case-marking feature from the verb. Thus, a verb with the morphological change (i.e., *eaten*) does not assign the accusative case to the internal argument of the verb in the direct object position (i.e., **eaten the fish*). This violates the Case Filter that assumes that overt noun phrases must be marked with cases. To avoid the violation of the Case Filter, the internal argument of the verb is moved to the specifier of the TP

(Tense Phrase) to receive the Nominative case. In the case of English, the *be*-verb must be inserted. The passive sentence has a corresponding active sentence. English and Japanese have this type of the passive sentence.

Japanese has another type of passive, which is termed *the indirect passive*. The following example modified from Johnson (2008, p. 365, 367) shows the direct passive and the indirect passive.

(2) a. Taro-ga otousan-ni shika-rare-ta.

Taro-NOM (his) father-by scold-PASSIVE-PAST

"Taro was scolded by his father."

b. imouto-ga otouto-ni kamera-o kowas-are-ta.

(my) sister-NOM (my) brother-by (her) camera-ACC break-PASSIVE-PAST

"My sister was upset because my brother broke her camera."

Sentence (2-a) is an example of the direct passive, and sentence (2-b) is an example of the indirect passive. The verb in (2-b) contains the suffix for passivization, *–rare*, but the internal argument with the accusative case of the verb (i.e., *o*-marked NP) remains in the position of the object of the verb. According to Izumi and Lakshmanan (1998, p. 68), the passive morpheme *–rare* has its allomorph *–are* as in *kowas-are* (break-PASSIVE), *kak-are* (write-PASSIVE), and *yob-are* (call-PASSIVE). According to Abe (1999) and Izumi and Lakshmanan (1998), the passive morpheme *–rare* in the indirect passive in Japanese is assumed to be a lexical/thematic verb that assigns the theta role of the experiencer and to have the untensed IP (inflectional phrase) as the internal argument of *–rare*.

The Courses of Study (the official guideline of the curricula) issued by Ministry of Education, Culture, Sports, Science and Technology (2008, 2010) stats that direct passives must be taught at junior high schools; nevertheless, the Courses of Study for junior and senior high schools does not mention that the ungrammaticality of indirect passives should be taught. Thus, it is not clear whether Japanese learners of English learn that indirect passives are not allowed in English.

Although Izumi and Lakshmanan (1998) used intransitive verbs such as *come* as in *I was come by my friend late last night,* the verbs used in the tests and the treatments in this study only consisted of transitive verbs that do not take indirect objects. The verbs used in the treatment tasks were *steal, eat, break, scold,* and *lose.* The oral production test elicited two new verbs (*drink, read*) as well as two of the verbs that were used in the treatment tasks (*lose, break*). The oral imitation test and the grammaticality judgment test tested two new verbs (*smoke* and *kill* for the oral imitation test; *read* and *kick* for the grammaticality judgment test) as well as three of the verbs that were used in the treatment tasks (*eat, break,* and *steal* for the oral imitation test; *scold, lose*, and *break* for the grammaticality judgment test). The test sentences were run through Vocabprofiler 2.7 (Cobb, 2006) to check their frequency levels; all the words except for 15 words were within the highest frequency 2,000 words of English. One word, *computer,* was in the Academic Word List (Coxhead, 2000), and 14 words—*bike, diary, cookies, hamburgers, Japanese, shogi, blackboard, mobile [phone],* *necklace, American, Chinese, French, wallet,* and *steak*—were in neither the high frequency word list nor the Academic Word List.

Reasons for Selecting Indirect Passives

The first reason that indirect passives was chosen is because it poses a learnability problem for Japanese learners of English. English allows direct passives only whereas Japanese allows both direct and indirect passives. Thus, once Japanese learners of English possess the rules overgeneralized from their L1, they have difficulty in learning that indirect passives are not allowed in English even though they are exposed to direct passives.

Second, Japanese learners of English tend to make persistent errors in the structures and are likely to accept or produce the indirect passive in English (e.g., Izumi & Lakshmanan, 1998; Suzuki, 1996). According to Izumi and Lakshmanan, it is predicted that Japanese learners of English transfer L1 rules to the L2 and judge both direct passives and indirect passives as acceptable in English. More specifically, Izumi and Lakshmanan hypothesized that Japanese learners of English initially treat the passive auxiliary *be* as a lexical-thematic verb like the passive morpheme *–rare* in Japanese and that learners think that the indirect passive is allowed in English. In order to check whether Japanese learners of English accept indirect passives in English, I examined the results of one item in the placement test administered to second-year university students of the Faculty of Education of University S in April, 2010. The placement test was carried out for the two-year required English course. The test included the following multiple-choice question.

[Question] Brian ( ) yesterday.

[Choices] 1. was read his diary by his mother

2. has his diary by his mother

3. was upset because his diary was read by his mother

4. made his diary read by his mother

The total number of students was about 290. Because 21 students did not agree to participate on the consent form, the responses from 269 students were analyzed. The results are summarized in Table 6. Of the 269 students, 36.1% chose the indirect passive, whereas 46.5% chose the direct passive and 14.5% chose a different structure. The results indicate that some Japanese university students overgeneralize the Japanese indirect passive structures to English.

Table 6. *Frequency and Rate of Students' Choice of Each Alternative (Passivization)*

	Frequency	Rate
1	97	36.1%
2	6	2.2%
3	125	46.5%
4	39	14.5%
No response	2	0.7%

Third, passivization requires high-level processing procedures such as the S-procedure in the processability theory (Pienemann, Di Biase, & Kawaguchi, 2005, p. 242). Therefore, compared with the adjective-ordering rules, the indirect passive is difficult in terms of processability within the learner interlanguage system. However, because the participants in Izumi and Lakshmanan (1998) and Suzuki (1996) produced direct and indirect passive sentences, the participants in this study, who resemble those of Izumi and Lakshmanan (1998) and Suzuki (1996) in terms of age, were not assumed to have great difficulty in processing passivization.

Instrumentation

Testing Tasks

Three instruments were used to measure the participants' learning the ungrammaticality of the overgeneralized rules: untimed grammaticality judgment tests, oral production tests, and elicited imitation tests. In order to avoid the influence of one test on another, more communicative tasks were given first, and the test that required participants to focus on the linguistic forms was administered later. Thus, the tests were presented in the following order: oral production, elicited imitation, and untimed grammaticality judgment tests.

These instruments were used to obtain (a) oral production, (b) imitation performance, and (c) untimed judgment. The first oral production and imitation performance measures were used as the dependent variables for the assessment of implicit knowledge of the target structures. These measurements were obtained in tasks in which the participants were under pressure to produce English in real time and were focused primarily on meaning rather than form. The untimed judgment measure was used as a dependent variable to assess explicit knowledge; because the participants were able to take time when making a judgment, they had an opportunity to access their explicit knowledge. All the tests were carried out on a computer screen.

Oral Production Task

The participants completed two oral production tasks (one for adjective-ordering rules and one for the indirect passive) that were administered to assess whether they can use interlanguage structures productively and avoid using interlanguage structures after the treatment. The task designed to elicit noun phrases containing adjectives was a spot-the-difference task in which the participants were shown two pictures and told that one was the original picture while the other was a slightly altered version of the picture (see Appendix H). They were required to find the differences from the original picture and to report those differences by using the phrase, "In picture 1, there is/are ---. But in picture 2, there is/are ---." The pictures were designed to elicit eight tokens of the target structures (*a short blue boat, a long red boat, a large yellow computer, a small blue computer, a low blue chair, a high gray chair, a small gray radio*, and *a large orange radio*). Before the task, an explanation of the testing procedures and a simple example task with two pictures were provided. It took the participants approximately 5 minutes to complete the task.

The task designed to elicit the indirect passive was a picture-cued story-making task (see Appendix I). The participants were given a series of pictures that depict a character's bad experiences and were required to make a story about the character. Before narrating the story, they were shown a series of pictures and were required to listen to an explanation about the pictures in their first language. The pictures were designed to elicit four tokens of the target structures. For this task, there was no training session. Three forms of the tasks differing in the order of pictures (A, B, and C) were created: Form A for the pretest, Form B for the immediate posttest, and Form C for the delayed posttest. Each task took about 5 minutes.

Elicited Imitation

I basically followed the procedures of the elicited imitation test described in Ellis (2005) and Erlam (2006). In order to make elicited imitation reconstructive in nature, (a) participants should be expected to process the language in real-time, (b) they should focus primarily on meaning rather than form, (c) there should be some delay between the provision of the test sentences and the repetition of the sentences, and (d) they should be required to repeat the sentences in correct English (i.e., to correct ungrammatical sentences).

This test consisted of 30 sentences (Appendix J): 10 tokens (five grammatical and five ungrammatical sentences) for each target structure and 10 distractor sentences (five grammatical and five ungrammatical sentences). For adjective ordering, grammatical items contained the structures with the size-color order of adjectives whereas the order was reversed (i.e., color-size) for ungrammatical items. For indirect passives, grammatical items consisted of the passive constructions with double-object verbs (*buy, give, show, teach*, and *tell*) in which indirect objects were moved to the subject position as in *David was bought a new coat last Sunday*; on the other hand, ungrammatical items contained indirect passives (e.g., *Mike was stolen his nice pen* and *David was smoked a cigarette by a lady*). The distractors focused on past tense forms of verbs. The sentences differed from those used in the grammaticality judgment tests and were ordered randomly.

The participants listened to the audiotaped sentences provided in real-time, in other words, under time pressure. The sentences were read slowly and clearly but naturally by a female American speaker of English. The participants were asked to identify the pictures that the sentences described (see Appendix K), as this would encourage them to focus primarily on meaning and resulted in some interval between the time of the stimulus presentation and that of imitation. Then they were asked to repeat the sentences orally in correct English so that they were aware of the correction condition. Four test items were given for the purpose of training before administering the targeted items.

Three forms of the test (A, B, and C) were created by changing the item order. Form A was administered as the pretest; Form B as the immediate posttest; and Form C as the delayed posttest. The participants' sentence imitation performances were audio-recorded. The time for the test was about 15 minutes. The recorded performances were transcribed and analyzed.

Untimed Grammaticality Judgment Tests

According to Ellis (1991), grammaticality judgment tests vary in terms of "whether they require the learner to (a) discriminate between well-formed and deviant sentences, (b) locate the errors in the deviant sentences, (c) correct the errors, or (d) provide grammatical descriptions of the errors" (p. 162). The grammaticality judgment test for this study was designed to require the participants to do the first task only, that is, to judge the grammaticality of a sentence. Because it was used as a measure of explicit knowledge, no time limit was assigned.

The untimed grammaticality judgment test consisted of 30 sentences (Appendix L): 10 sentences (five grammatical and five ungrammatical sentences) for each target structure and 10 distractor sentences (10 grammatical and 10 ungrammatical sentences). For adjective ordering, grammatical items contained structures with the size-color order of adjectives whereas the reversed order (i.e., color-size) was created for ungrammatical items. For grammatical items for passivization, the passives containing double-object verbs (*buy, give, show, teach*, and *tell*) with their indirect objects as the subjects were created as *Hideki was given his first watch as a Christmas present* whereas ungrammatical items consisted of indirect passives (e.g., *Taro was read his love letter by his mother* and *Brian was kicked his leg by the man in the train*). The distractor focused on past tense. The test sentences were ordered randomly.

The participants were presented with one test sentence in written form and a picture that the sentence describes and were asked to judge the grammaticality of the sentences on a four-point rating scale, including "the degree of confidence" (Gass, 1994, p. 310): 1 = *Incorrect (tadashikunai)*; 2 = *Somewhat incorrect (amari tadashikunai)*; 3 = *Somewhat correct (yaya tadashii)*; and 4 = *Correct (tadashii)*. In order to ensure that participants paid attention to every word while reading, a recording of the sentences read by one American speaker of English (female) was played. Thus, the medium of presentation was both written and aural. The participants answered the questions at their own pace. That is, even after the recording ended, they could take their time for their judgments. They were told that they must not return to earlier items. The untimed grammaticality judgment test took about 10 minutes.

Three forms of the test whose orders of the items were different (A, B, and C) were prepared: Form A for the pretest, Form B for the immediate posttest, and Form C for the delayed posttest. Each test contained an explanation of the testing procedures and three examples (one grammatical sentence and two ungrammatical sentences) for the purpose of training.

Following Ellis's (1991, pp. 170-171) recommendations concerning test design, Table 7 summarizes the design features of the grammaticality judgment test used for this study.

Table 7. *Designs of the Grammaticality Judgment Test for This Study*

Features	The study
Task features	
Order of presentation	Randomized
Distractor items	Yes
Medium of presentation	Written and aural
Controlling linguistic complexity	No
Contextualization	With pictures
Procedure features	
Number of responses requested	Ranking (4-point scale)
Immediacy of response	No time limit
Timed response	No
Familiarization with task requirements	Yes

Treatment Tasks

The participants met individually with me for the treatment sessions. The treatment sessions consisted of three types of information-gap tasks (an object placement task, a picture identification task, and a picture description task) and two types of tasks (a story-making task at the beginning of the session and a spot-the-difference task at the end of the session) as distractor tasks.

In the object placement task, the participants in all the groups received positive evidence of the target structures; in other words, participants listened to recorded directions to place items with different sizes and colors and name tags and were told to place the items and name tags at the proper points (Appendix M). First, participants were given the small board with an outdoor scene including figures who had adverse experiences, 20 pieces of picture cards with different sizes and colors, and five name tags. They were told that they would hear the recorded directions once in English from a speaker and were instructed to choose appropriate items and names and to place them on the board as soon and accurately as possible. When they were ready for the next direction, they were asked to signal me by saying, "OK." Then, I played the next recorded sounds. There were ten directions: five for adjective-ordering rules (e.g., *Please put a big green bird under the clouds* and *Please put a tall black tree by the street*) and five for passivization (e.g., *Ken was upset because his wallet was stolen at the station* and *Taro was upset because his sister ate his cake*). The ten sentences were read by a female American speaker of English. The sentences were mixed so that the participants listened to the adjective order and the passive in turn. The following shows the interaction during the task derived from the data of a participant (ID # 1013) from the POS Group.

Researcher:	[The researcher played the direction.]
(A voice)	Please put a big green bird under the clouds.
Learner:	[The learner chose the item and placed it on the board.] OK.
Researcher:	[The researcher played the direction.]
(A voice)	Ken was upset because his wallet was stolen at the station.
Learner:	[The learner chose the name and placed it on the board.] OK.

Through this activity, the participants were exposed to five tokens of positive evidence about the target structures, without any salience, for the REC Group, the REC+ Group, and the POS Group, and with salience by stress and intonation for the SAL Group.

The picture identification task was used for the provision of positive evidence for the POS and SAL groups. For these two groups, the task was performed in English. The participants and I held a set of ten cards consisting of four pictures on each. I described one of four similar pictures in English so that the participants could identify that picture (Appendix N). For example, I said, "In this picture, I see a little white flower." Then, the participants selected the picture of a little white flower among the four pictures with different colors and sizes. The four pictures on a card were numbered from one to four so that the participants gave their answer by saying the number of the picture like "No. 2." I gave feedback about the correctness of their answer and went on to the next card. The following example came from a participant (ID # 1013) from the POS Group.

Researcher:	In this picture, I see a little white flower.
Learner:	Number 2.
Researcher:	OK. *Tsugi ikimasu ne* [Let's move on to the next card.] Yumi was unhappy because her nice cup was broken.
Learner:	1.
Researcher:	OK. In this picture, I see a big black bag.
Learner:	2.
Researcher:	OK.

The activity was designed so that the participants had five chances to comprehend each form. The cards were arranged so that the participants listened to the two target structures in turn. On the pictures on the five cards for passivization, the names and the information about what had happened (e.g., *kowasareta* meaning "was broken" and *nusumareta* meaning "was stolen") were written in Japanese in order to facilitate the participants' interpretations of pictures.

For the REC and REC+ Groups, the picture identification task was performed in Japanese, during which the participants were provided the color-size-noun sequence and the indirect passives in Japanese (e.g., *Korewa shirokute chiisana hana desu* meaning "This is a white small flower" and *Yumiwa kanojono sutekina kappuo kowasaremashita* meaning "Yumi was broken her nice cup"). There were two reasons for this. First, the picture identification task was carried out in Japanese in order to control the opportunities for the participants to think about the pictures before they were required to describe them in the following picture description task. The participants in the POS and SAL groups had the picture identification task that was performed in English. Thus, the order of the two tasks was considered to be helpful for the participants in the POS and SAL groups to think about the pictures. It was expected that the picture identification task that was performed in Japanese could give the participants in the REC and REC+

Groups the same chances to think about the pictures that were used in the following picture description task, without providing the positive evidence in English. The second reason was that giving the color-size-noun sequence and the indirect passives in Japanese would increase the chances that the participants in the REC and REC+ groups would produce the sequence and the indirect passives in English in the picture description task. The participants were not required to describe that picture immediately after they listened to the Japanese description of each picture; therefore, the participants were considered to produce the color-size-noun sequence or the indirect passives in English in the picture description task only when they had the interlanguage grammar of English that allowed the color-size-noun sequence or the indirect passives.

The picture description task was used for the provision of recasts for the two recast groups (the REC and the REC+ Groups) and for the provision of output opportunities for the POS and SAL Groups. In this task, the same set of ten cards as the picture identification task was used. Participants were asked to choose one of the four pictures on each card and describe the picture in one English sentence so that I could identify that picture. The participants in the REC and REC+ Groups had not only output opportunities but also opportunities to receive recasts from me: On the other hand, the participants in the POS and SAL Groups had output opportunities only and did not receive any form of negative feedback including recasts. The following example derived from a participant (ID # 1013) from the POS Group shows the interaction during the task.

Learner:	I see a big pink flower.
Researcher:	Number 3.
Learner:	Yes. [The participant looked at the next card.] Mike was upset because his nice cup was broken.
Researcher:	Number 3.
Learner:	Yes. [The participant looked at the next card.] I see a big white bag.
Researcher:	Ah, number 1.
Learner:	Yes.

At the beginning of the session, all the four groups performed a story-making task as a distractor task. In this task, the participants and the researcher had the same set of ten pictures depicting actions. The participant was asked to make a story about "What did Yuki [a girl's name] do yesterday?" based on the ten pictures so that the researcher could put the pictures in order. In addition, at the end of the session, all four groups performed a spot-the-difference task as a distractor task. The participant was given two pictures that describe a scene at a restaurant and was asked to report as many differences between the two pictures as possible in two minutes. Both distractor tasks were not designed to elicit the target structures of this study.

Participants in the REC and REC+ Groups received at most five tokens of recasts for each structure. The number of recasts was decided for the following reasons. First, I hoped to minimize the number of recasts so that the findings of this study can be generalized to natural interactions in communicative language classrooms. Second, I checked the number of recasts among the previous experimental studies on the effectiveness of recasts and found that the number of recasts provided to individual participants ranged from six (Long, Inagaki, & Ortega, 1998) to 24 (Carroll & Swain, 1993). Long, Inagaki, and Ortega (1998) carried out two studies whose target structures were adjective-ordering rules in Japanese and adverb placement rules for English-speaking learners. They provided six instances of recasts and found the effectiveness of recasts for Spanish adverb placement rules. For this study, the participants in the REC and REC+ Groups received five tokens of positive evidence for each structure as well as recasts. Taken together, the number of recasts was five.

Procedures

Treatment

The REC Group performed the object placement task, the picture identification task (L1), and the picture description task with the provision of recasts. In the object placement task, the participants in the REC Group listened to instructions containing the target structures and received positive evidence about the target structures. In the picture identification task (L1), they listened to descriptions in Japanese about the pictures and were told to identify the pictures. In the picture description task, the participants were told to describe one picture in English among four similar pictures so that I could identify that picture. The participants received recasts and non-corrective repetitions in the form of confirmation checks: In other words, I repeated every grammatical utterance or reformulated every erroneous utterance.

Learner:	This is a *big white bag*. [grammatical]
Researcher:	This is a *big white bag*? [non-corrective repetition]
	[The researcher will identify the picture.]
Learner:	There is a *black short T-shir*t. [erroneous]
Researcher:	There is a *short black T-shirt*? [recast]
	[The researcher will identify the picture.]

Non-corrective repetition was provided to make the amount of positive evidence and output opportunities comparable to those of the other groups.

The participants in the POS Group had the same number of chances to listen to the target structures and to produce the target structures as the participants in the REC group. They performed three tasks: the object placement task, the picture identification task (L2), and the picture description task (without provision of recasts). The object-placement task was carried out in the same way as the REC Group. In the picture identification task, the participants listened to my description of the pictures in English and identified those pictures. In the picture description task, the participants played the role of the describer. What differed from the REC Group was the non-provision of feedback in the form of non-corrective repetition or recasts.

Learner:	This is a *big white bag*. [grammatical]
Researcher:	OK. [The researcher will identify the picture.]
Learner:	There is a girl with a *black short T-shirt*. [erroneous]
Researcher:	I see. [The researcher will identify the picture.]

Thus, the participants were provided with the same number of output opportunities without receiving any positive or negative evidence as the REC Group.

The participants in the REC+ Group performed the picture identification task (L1), the picture description task (with provision of recasts), and the object placement task), that is, in a different order from that of the REC Group. The provision of feedback was carried out in the same way as for the REC Group. For the REC+ Group, the object placement task allowed the participants, who had received recasts on the picture description task, to be exposed to further input containing the target structures. Thus, the numbers of recasts, positive evidence, and output opportunities were the same for the REC and REC+ Groups, but the timing of the provision of positive evidence differed: before the provision of recasts for the REC Group and after the provision of recasts for the REC+ Group.

Lastly, the participants in the SAL Group performed the three tasks in the same way as the POS Group. The difference was the manner of the provision of positive evidence. For the SAL Group, the target structures were provided with the enhanced salience by stress and intonation.

Table 8 shows the treatments for the four groups. In total, the treatment session lasted approximately 30 minutes.

Table 8. *Four Treatment Procedures*

Group	Treatment tasks				
POS	Object placement task (input)	→	Picture identification task (input)	→	Picture description task (output)
REC	Object placement task (input)	→	Picture identification task (L1)	→	Picture description task (output with recasts/non-corrective repetition)
REC+	Picture identification task (L1)	→	Picture description task (output with recasts/non-corrective repetition)	→	Object placement task (input)
SAL	Object placement task (input with enhanced salience)	→	Picture identification task (input with enhanced salience)	→	Picture description task (output)

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Operationalizing Recasts

As discussed in the Definitional Issues section in Chapter 2, recasts have been defined in various ways. For this study, a recast is operationalized as a reformulation of all of an incorrect utterance with rising intonation (that is, in the form of a confirmation check) as in the following examples from data in the current study:

Example from the Picture Identification Task (1026, REC Group)
(a) Adjective ordering

Learner:	There is green small house. [error]
Researcher:	There is a small green house? [recast]
Learner:	Yeah.
Researcher:	Number 3.

(b) Indirect passives

Learner:	Mike is broken his cup. [error]
Researcher:	Mike is sad because his cup was broken? [recast]
Learner:	Yeah.
Researcher:	Number 3.

Stress and emphasis were not added to recasts; recasts were provided naturally.

Recasts for adjective ordering and indirect passives were the same in terms of their corrective nature, segmentation/expansion, and intonation. However, the length of the recasts and the number of changes varied according to the erroneous nature of the preceding utterances. The number of changes in the recasts for adjective ordering tended to be fewer than that of indirect passives, and the length of recasts for adjective ordering was shorter than that of indirect passives.

Lastly, recasts were given frequently in a short of period of time in this experiment. In order to control the amount of input and output opportunities, recasts or non-corrective repetitions were given to every utterance of L2 learners during the treatment tasks. Thus, the conversational interactions observed in the study were not the same as natural ones in terms of frequencies of recasts.

Consistency Across Groups

The following factors were controlled: opportunities for production, opportunities for receiving positive evidence in the form of recasts for the REC and REC+ Groups and non-contingent positive evidence for the POS and SAL Groups, and no opportunities for modified output. Table 9 indicates the consistencies and differences across the groups.

Table 9. *Treatment Group Conditions*

	POS	REC	REC+	SAL
Production opportunities	+	+	+	+
Positive evidence	+ (2 tasks)	+	+	+
Positive evidence through recasts/non-corrective repetition	-	+	+	-
Enhanced salience	-	-	-	+
Modified output opportunities	-	-	-	-
Order	Input Input Output	Input Output/input*	Output/input* Input	Input Input Output

Note. * output/input = recast condition. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

The POS and SAL Groups differed in that the latter was provided with positive evidence with enhanced salience. The REC and REC+ Groups differed in the position of the provision of positive evidence: The REC Group carried out the input task before the task in which recasts were provided, whereas the REC+ Group carried out the input task after recasts were provided, which gave that group additional input.

The order of the production activities for the POS and SAL Groups differed from the conditions used by Leeman (2003) in which the production activity was followed by the provision of positive evidence. According to Swain's (1985, 1993, 1995) output hypothesis, production in itself is hypothesized to bring about noticing by the learners and facilitate their motivation to search for the necessary information in the further input (see Sakai, 2004; Swain, 1998; Swain & Lapkin, 1998 for empirical support). To reduce the effects of production, the output activities took place after the input activities.

Analysis

The recorded data on the pretest, posttest, and delayed posttest for the oral production test, elicited imitation test, and grammaticality judgment test were transcribed by two assistants; the transcription took about 170 hours. I checked all the transcriptions. Although I found several errors (e.g., spelling mistakes), they did not influence the analyses. I coded and analyzed the rechecked transcription as stated below.

Data Coding

Oral production test. For adjective ordering, out of the eight targeted tokens, each token of the participants' production on oral production tests was coded as *correct* (target-like use), *incorrect* (non-target use), and *non-application*. The following is an example of the transcribed performance on the spot-the-difference task as a pretest.

> Example from the Spot-the Difference Task (1028, POS group)
> The in picture 1, (1) there is *a red curtain.* But, in picture 2, there is *a blue curtain*. Eh, in picture 1, (2) out of the window (1) Ah, (1) there is a blue boat *blue short boat* (3) out of the window. But, (1) in picture 2, there is *a long red boat* out of the window. In picture 1, there is *a yellow computer.* But, in picture 2, there is *a blue computer*. And the (1) size of computer is different. (2) In picture 1, the computer is ah, *the computer of picture 1 is bigger* than picture the computer of picture 2. Eh, in picture 1, *the (2) radio (1) is (1) black and small*. But, in picture 2, *the radio is orange and big*.
>
> *Note*. The parenthesized number indicates the duration of a pause in seconds. The italics indicate the parts that contain adjective ordering.

The participant (ID # 1028 from the POS Group) produced two correct instances of noun phrases containing the two types of adjectives (size and color) as premodifiers (*blue short boat* and *a long red boat*). The two phrases were coded as incorrect for *blue short boat* (color-size-noun) and correct for *a long red boat* (size-color-noun). I analyzed the adjective combinations premodifying nouns only; therefore, I coded phrases containing adjectives used predicatively as in *radio is black and small* and noun phrases premodified by only one type of adjectives, as in *a red curtain* and *a blue computer*, as a non-application of adjective ordering.

For passivization, four tokens of the participants' production were coded as *correct, incorrect, indirect passive*, and *non-application*. First, the category *correct* refers to the correct use of passive voice (e.g., *His cap was lost*), the use of active voice (e.g., *Brian's brother lost his cap*), or the use of the *have + object + past participle* structure (e.g., *Brian had his cap lost by his brother*). Second, the category *incorrect* refers to the wrong description of a picture like *Brian lost his black cap* (it was Brian's brother who lost his cap). Third, the category *indirect passivization* was used for tokens that were incorrect in terms of English passivization but whose equivalent structure in Japanese was acceptable (e.g., *Brian was stolen his favorite cap*). Lastly, non-application was used for the case in which target expressions were not observed.

I coded all the data. After a one-week interval, I recoded the data without referring to the first coding to calculate intra-coder reliability. The agreement reached 99.3% for adjective ordering and 98.8% for passivization. For both structures, the disagreement was caused not by the ambiguity of judgment, but due to my carelessness. Thus, I rechecked the disagreement codes against the data and made final decision.

Elicited imitation tests. On the elicited imitation tests, two types of responses were obtained: picture selections and imitations. For this study, I took the accuracy of picture selections into consideration. That is, correct performance of imitation with a wrong selection of a picture was not considered correct. Such a case might show that the participant directed attention to linguistic form rather than to meaning. To ensure the reconstructive nature of imitation on the elicited imitation tests, I first coded the picture selections dichotomously as either correct or incorrect. In the case of the correct picture selections, the imitations were coded as stated below.

In principle, the imitations were coded into three categories: *accept* (the verbatim imitation of the target part of a sentence), *reject* (the changes of the target part), and *other.* Performances coded as *accept* were considered correct for a grammatical item or incorrect for an ungrammatical item. In contrast, performances coded as *reject* were regarded as incorrect for a grammatical item or correct for an ungrammatical item. The category *other* contains non-imitation or partial imitation of the target part of a heard sentence, so it was difficult to judge whether the participants attempted to accept or reject the target part. Regarding passivization, the responses with changes in the target parts of test items were coded as *reject*, whether the meaning was correct or incorrect. This approach was taken to examine the degree to which the participants learned that some forms are ungrammatical in the L2, not the degree to which they learned the grammatical structures. Changing the target part of an ungrammatical item might indicate knowledge that the presented structure is ungrammatical; on the other hand, even changing the target part of a grammatical item into a correct structure (e.g., "Jane was told the story by Mike." to "Mike told his story to Jane.") implies that the participant believes that the structure is not acceptable in the L2.

Table 10 shows examples of codes from the present study. The test items for adjective ordering and indirect passive were "There is an orange short bus in the park" and "David was smoked a cigarette by a lady" respectively. Both sentences are ungrammatical. First, the imitations coded as *accept* (i.e., *incorrect*) are "There is orange short bus in the street" for adjective ordering and "David was smoken cigarette by lady" and "David is smoked a cigarette by a lady" for indirect passive. This example for adjective ordering shows the participant's acceptance of the color-size-noun sequence. For indirect passive, the two examples show that the participants accepted the "David + be + past participle" sequence although the participants did not imitat the test item perfectly: The form of the past participle was an interlanguage form (*smoke)* in the first example whereas the tense was wrong in the second example. The imitations coded as *reject* (i.e., *correct*) are "There is very short orange bus on the street" for adjective ordering and "David was David don't like … her smoking," "David was smoking cigarette by a lady," and "A David, ah, smokes by lady" for indirect passive. The example for adjective ordering contains the correct sequence of adjectives (i.e., the size-color-noun sequence, *short orange bus*). The three examples show that the participants avoided the expression "Dave was smoked a cigarette" and changed the expressions.

The latter two examples were categorized as reject (i.e., correct) although they did not contain the correct meanings of the test item. As stated above, this was because changing the expressions of indirect passive was considered to indicate that participants reject indirect passives. The imitations categorized as *other* are “There is a orange bus on the street” and “There is an short bus on the tree” for adjective ordering and “David was … ciga lady” and “David was cigarette by a lady” for indirect passive. The examples for adjective ordering contained only one adjective and thus were impossible to judge whether participants avoid or accept the color-size-noun sequence. In the examples for indirect passive, the target part (i.e., *David was smoked*) was omitted; thus, it was difficult to judge whether participants accepted indirect passives but omitted some words or they rejected indirect passives. As the examples show, errors other than the target forms (e.g., article, tense) were not considered in the scoring.

Table 10. *Examples for Coding of Imitation Responses to Ungrammatical Items on the Elicited Imitation Tests*

	Adjective ordering	Indirect passive
Test item	*There is an *orange short* bus in the park.	*David *was smoked* a cigarette by a lady.
Code		
Accept (incorrect)	There is *orange short* bus in the street.	David *was smoken* cigarette by lady. David *is smoked* a cigarette by a lady.
Reject (correct)	There is very *short orange* bus on the street.	David was David don't like … her smoking. David was smoking cigarette by a lady. A David, ah, smokes by lady.
Other	There is a *orange* bus on the street. There is an *short* bus on the tree.	David was … ciga lady. David was cigarette by a lady.

Scoring the participants’ imitations for adjective ordering was done relatively mechanically and was easier than scoring the imitations for indirect passives. For indirect passives, it was difficult to judge whether participants accepted indirect passives. Because errors other than the target forms were not considered, imitations that are not verbatim repetitions of the test items can be coded as *accept*. I coded all the data. Another rater coded 17% of the data set to establish inter-rater reliability. The agreement reached 93.0%.

Grammaticality judgment tests. I analyzed (a) the data of the four scales as they were, that is, including the degree of confidence, and (b) the dichotomous data obtained by collapsing the four possible responses into two categories, correct and incorrect. For the former, I calculated the mean scores for each participant. The score ranges from 1 to 4. The higher score indicates that the participant tended to accept the sentence as grammatical, while the lower score indicates that the participant tended to reject the sentence as ungrammatical. In the latter case, I summed the numbers of correct responses; thus, the score for each participant ranged from 0 to 5. For example, a score of 3 indicates that the participant correctly judged three out of the five items.

Table 11 shows the reliability coefficients for adjective ordering and passivization. The reliability coefficients were relatively low for grammatical items: .59 to .79 for adjective ordering and .34 to .54 for passivization. Thus, caution is needed when interpreting the results for the grammatical items.

Table 11. *Cronbach's Alpha Estimates for the Grammaticality Judgment Tests*

		Ungrammatical items (k = 5)			Grammatical items (k = 5)		
Target form	*n*	Pretest	Posttest	Delayed posttest	Pretest	Posttest	Delayed posttest
ADJ	75	.64	.86	.88	.62	.59	.79
IDP	90	.60	.81	.78	.34	.53	.54

Note. ADJ = adjective ordering; IDP = indirect passives.

Data Analysis

For each test, I analyzed both group results and individual results. For group results, I focused on the participants' performance of ungrammaticality for the two target forms. In other words, I examined whether the production of ungrammatical structures on the oral production tests changed across the groups and whether the groups differed in their performance on ungrammatical items on the elicited tests and the timed grammaticality judgment tests.

First, the descriptive statistics were calculated for the dependent variables for the four groups. Then, I checked whether the distributions deviated from normality by examining the values of skewness and kurtosis. The values of z-skewness and z-kurtosis that were obtained by dividing skewness and kurtosis by their standard errors were considered to be significant at $p < .05$ if their absolute values are greater than 1.96 (Field, 2009, p. 139). Kolmogorov-Smirnov tests were also performed to test the normality of each variable.

As shown in the Results section, the distributions of several variables deviated from normality, so nonparametric statistical procedures were used. As to the first research question, a Friedman test was performed to examine whether the POS Group changed significantly between the three tests. The alpha was set at .05. Regarding the other four research questions, a difference score was computed between two of the three tests and three Kruskal-Wallis tests were performed on the difference scores for the pretest-posttest, the pretest-delayed posttest, and the posttest-delayed posttest periods to determine whether the groups differed in terms of the changes in the scores. Alpha was set at .017 with a Bonferroni adjustment (.05/3 periods). As a post hoc test, the Mann-Whitney test was performed to make focused comparisons in accordance with the research questions/hypotheses: the POS Group vs. the REC Group (for research question 2), the REC Group vs. the REC+ Group (for research question 3), the POS Group vs. the SAL Group (for research question 4), and the REC Group vs. the SAL Group (for research question 5). As Field (2009) suggested, effect sizes were not calculated for the Kruskal-Wallis test; but effect sizes as measured by Pearson's correlation coefficient r were calculated for focused comparisons (p. 570). The criteria were .10 for a small effect, .30 for a medium effect, and .50 for a large effect (Cohen, 1988, pp. 79-80; Field, 2009, p. 57).

For this study, not only the group data but also data from individuals were analyzed as in Eckman (1994). Eckman pointed out that it is necessary to examine systematicity in one individual's responses. In order to consider the participants' systematic production or judgment of the target structures, a cut-off point was set. For the oral production tests, I regarded participants who produced no incorrect structures plus at least one correct structure as those who avoided incorrect structures systematically. For the elicited imitation tests and the grammaticality judgment tests, I considered the participants who responded to three or more of the five items (i.e., 60%)

in the same way as showing systematic responses. For individual results, I analyzed how the performance by those participants who had produced or accepted both correct structures and incorrect structures changed on the posttest and the delayed posttest.

CHAPTER 4

RESULTS

Preliminary Analyses

The preliminary analyses involved examining the number of recasts provided in the treatment tasks. Table 12 shows the number of participants according to the number of recasts they received. For adjective ordering, for example, ten of the REC Group members received one recast, six received two recasts, two received three recasts, one received four recasts, and one received five recasts.

Table 12. *The Number of Participants According to the Number of Recasts Provided on the Treatment Tasks*

		Number of recasts					
Group	*n*	1	2	3	4	5	*M (SD)*
ADJ							
REC	20	10	6	2	1	1	1.85 (1.14)
REC+	16	6	6	2	1	1	2.06 (1.18)
IDP							
REC	22	4	4	2	11	1	3.05 (1.29)
REC+	27	5	5	4	10	3	3.04 (1.34)

Note. ADJ = adjective ordering; IDP = indirect passives; REC = REC Group; REC+ = REC+ Group.

To test the difference between the distributions of the number of recasts provided on the treatment tasks between the REC and REC+ Groups, Fisher's exact tests were performed on the two-by-five (group by number) contingency tables, instead of Chi-square tests, because some of the cells contained expected values of less than five. The tests show that there were no statistically significant differences between the two groups for adjective ordering (p = .970) or for indirect passives (p = .889). Thus, a statistically significant difference between the two groups was not found for each structure.

According to Table 12, the number of recasts differed between the two structures in each group, that is, participants received more recasts targeting indirect passives than targeting adjective ordering. I checked whether each group differed in the number of recasts they received between the two structures. Fisher's tests showed that the difference between the two structures was not significant for the REC+ Group (p = .132), but it was statistically significant for the REC Group (p = .009). The effects of the different numbers of recasts are discussed in the Discussion Chapter.

Adjective Ordering

In the following sections, I report the results of the three tests (oral production tests, elicited imitation tests, and untimed grammaticality judgment tests) for adjective ordering.

Oral Production Tests

Group results. Table 13 shows the descriptive statistics for the frequencies of incorrect adjective order on the oral production tests (see also Figure 1). For the REC Group, the mean score of the use of incorrect adjective order (color-size) increased from the pretest (0.30) to the posttest (1.10), but the mean score on the delayed posttest decreased (0.60). The mean score of the REC+ and SAL Groups increased from the pretest (0.44 and 0.42) through the posttest (0.75 and 1.21) to the delayed test (1.38 and 1.58). The POS Group displayed no marked changes between the tests.

Table 13. *Descriptive Statistics of the Frequencies of Incorrect Adjective Order on the Oral Production Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pretest	0.40	0.94	3.27	0.51	12.06	0.99
Posttest	0.75	1.41	2.00	0.51	3.50	0.99
Delayed	0.85	1.69	2.06	0.51	3.77	0.99
REC (*n* = 20)						
Pretest	0.30	0.73	3.02	0.51	10.00	0.99
Posttest	1.10	1.86	2.07	0.51	3.58	0.99
Delayed	0.60	1.19	1.94	0.51	2.87	0.99
REC+ (*n* = 16)						
Pretest	0.44	0.96	1.99	0.56	2.70	1.09
Posttest	0.75	1.58	3.05	0.56	10.30	1.09
Delayed	1.38	1.75	0.97	0.56	-0.46	1.09
SAL (*n* = 19)						
Pretest	0.42	0.77	2.35	0.52	6.48	1.01
Posttest	1.21	1.27	0.64	0.52	-0.61	1.01
Delayed	1.58	1.84	0.88	0.52	-0.02	1.01

Note. The highest possible score is 8. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

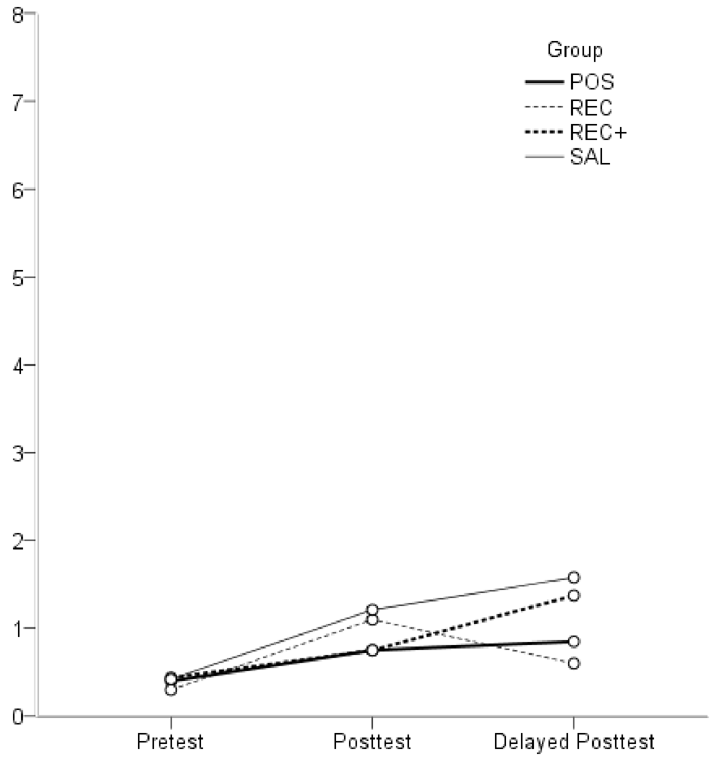


Figure 1. Mean frequencies of incorrect adjective order on the oral production tests. The highest possible score is 8.

As shown in Table 13, the assumption of normality of the distribution of each level of each test was not observed in terms of skewness and kurtosis. Therefore, as stated in the Data Analysis section, first a nonparametric Friedman test was performed to examine whether the POS Group changed between the tests. Next, statistical tests to investigate whether the difference scores between the tests differed between groups were performed.

The participants in the POS Group produced a mean of 0.40 instances of incorrect adjective order on the pretest. The mean score increased slightly to 0.75 on the posttest and showed a slight gain from the posttest to the delayed posttest (0.85). The Friedman test indicated that there were no statistical differences between the test results: $\chi^2(2) = 0.60$, $p = .755$. In other words, the POS Group did not show statistically significant changes between the tests.

Table 14 presents the descriptive statistics of the difference scores between two of the three tests on the oral production tests. The distribution of each variable deviated from normality in terms of skewness and kurtosis. Therefore, to compare the groups on the three tests, three Kruskal-Wallis tests were performed. The alpha level was set at .017 (.05/3 periods) with a Bonferroni adjustment. The tests showed no significant differences between the groups: $\chi^2(3) = 4.24$, $p = .237$ for the pretest-posttest period; $\chi^2(3) = 4.07$, $p = .254$ for the pretest-delayed posttest period; and $\chi^2(3) = 2.83$, $p = .418$ for the posttest-delayed posttest period. Furthermore, effect sizes for all the comparisons between the groups were very small (i.e., less than .10) or small (i.e., less than .30; see Appendix O). Thus, although Table 14 showed that the REC Group decreased the use of incorrect adjective order from the posttest to the delayed posttest (the difference score = -0.50), whereas the other groups increased or retained the use of incorrect adjective order for that period (see also Figure 1), the statistical analysis did not support that the groups differed significantly.

Table 14. *Descriptive Statistics of Difference Scores on the Oral Production Tests (Adjective Ordering)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pre-Post	0.35	1.63	1.15	0.51	3.24	0.99
Pre-Delayed	0.45	1.40	2.31	0.51	5.75	0.99
Post-Delayed	0.10	1.89	1.66	0.51	4.57	0.99
REC (*n* = 20)						
Pre-Post	0.80	1.54	2.28	0.51	6.34	0.99
Pre-Delayed	0.30	0.98	1.56	0.51	2.44	0.99
Post-Delayed	-0.50	1.50	-1.49	0.51	3.57	0.99
REC+ (*n* = 16)						
Pre-Post	0.31	1.54	0.15	0.56	2.41	1.09
Pre-Delayed	0.94	1.73	0.99	0.56	1.23	1.09
Post-Delayed	0.63	1.86	0.42	0.56	0.33	1.09
SAL (*n* = 19)						
Pre-Post	0.79	1.32	-1.04	0.52	2.74	1.01
Pre-Delayed	1.16	1.89	1.23	0.52	0.96	1.01
Post-Delayed	0.37	2.19	0.92	0.52	0.76	1.01

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Individual results. For the individual analysis, individual performances on the pretest of the production task were examined to check how many participants produced both correct and incorrect adjective orders, a pattern that might be due to the influence of the L1. Table 15 shows the participants who produced both the correct order (size-color) and the incorrect order (color-size): There were three out of 20 in the POS Group, two out of 20 in the REC Group, one out of 16 in the REC+ Group, and four out of 19 in the SAL Group. Fisher's exact test shows no statistically significant difference among the groups ($p = .656$) (see Table A in Appendix P for the contingency table).

Table 15. *Individual Analysis of Performance on the Oral Production Tests (Adjective Ordering)*

		Pretest			Posttest			Delayed posttest		
Participant	Effects	COR	INC	Other	COR	INC	Other	COR	INC	Other
POS (*n* = 20)										
1028	Delayed	1	1	6	7	1	0	8	0	0
1096	Short	3	1	4	1	0	7	1	6	1
1046		1	1	6	0	0	8	0	0	8
REC (*n* = 20)										
1083	Retained	1	1	6	1	0	7	8	0	0
1093	Delayed	1	1	6	0	1	7	5	0	3
REC+ (*n* = 16)										
1012		2	2	4	2	6	0	2	3	3
SAL (*n* = 19)										
1107	Delayed	1	1	6	6	2	0	4	0	4
1074	Short	3	3	2	1	0	7	5	3	0
1040		7	1	0	5	3	0	5	3	0
1108		6	1	1	4	4	0	6	2	0

Note. COR = correct order (size-color); INC = incorrect order (color-size); Other = instance without a combination of size and color, that is, no use of adjectives or use of one adjective. The shadowed cells indicate no-use of the incorrect order (color-size). Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the column of Effects refers to the participant who performed better on the posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Of the 10 participants, one (1083) in the REC Group did not produce the incorrect order on the posttest and retained no-use of the incorrect order on the delayed posttest (see the label Retained in the column of Effects of Table 15). Three participants (1028 in the POS Group, 1093 in the REC Group, and 1107 in the SAL group) did not produce the incorrect order on the delayed posttest. In other words, delayed effects were observed for these three participants. Short-term effects were found for participant 1096 in the POS Group and participant 1074 in the SAL Group. That is, these two participants did not use the incorrect order on the posttest but both produced the correct and incorrect orders on the delayed posttest.

A Fisher's exact test was performed on the number of participants who did not produce incorrect adjective order on the posttest or the delayed posttest (see the label Retained in Table 15. See Table B in Appendix P for the contingency table). The test showed no statistically significant difference among the groups (p = .300).

Furthermore, another Fisher's exact test was performed on the contingency table that contained the participants who showed delayed effects (see Table C in Appendix P); however, no statistically significant difference was observed (p = .429).

Elicited Imitation Tests

Group results. As stated in the Data Coding section, two types of data were obtained from the elicited imitation tests: picture selection and oral imitation. Imitation performance was analyzed with the correct selection of a picture. Therefore, before analyzing the imitation performance, I first checked whether the participants' selection of pictures for the ungrammatical items of the elicited imitation tests differed across the groups.

Picture selection. Table 16 shows the descriptive statistics for picture selection (see also Figure 2). In general, the scores were high, ranging from 4.42 to 4.90; the standard deviations were small, ranging from 0.31 to 0.72. Thus, the participants in the four groups performed well in the selection of pictures for the ungrammatical items (the color-size order) on the elicited imitation tests. This suggests that the participants focused on the meanings of the items and understood the sentences well.

Table 16. *Descriptive Statistics of Correct Picture Selection for Ungrammatical Items (Color-Size) on the Elicited Imitation Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pretest	4.70	0.47	-0.95	0.51	-1.24	0.99
Posttest	4.75	0.44	-1.25	0.51	-0.50	0.99
Delayed	4.85	0.37	-2.12	0.51	2.78	0.99
REC (*n* = 20)						
Pretest	4.65	0.67	-1.78	0.51	2.02	0.99
Posttest	4.90	0.31	-2.89	0.51	7.04	0.99
Delayed	4.90	0.31	-2.89	0.51	7.04	0.99
REC+ (*n* = 16)						
Pretest	4.62	0.72	-1.73	0.56	1.70	1.09
Posttest	4.75	0.45	-1.28	0.56	-0.44	1.09
Delayed	4.75	0.58	-2.38	0.56	5.31	1.09
SAL (*n* = 19)						
Pretest	4.42	0.61	-0.50	0.52	-0.50	1.01
Posttest	4.79	0.54	-2.66	0.52	6.88	1.01
Delayed	4.68	0.67	-2.00	0.52	2.81	1.01

Note. The highest possible score is 5. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

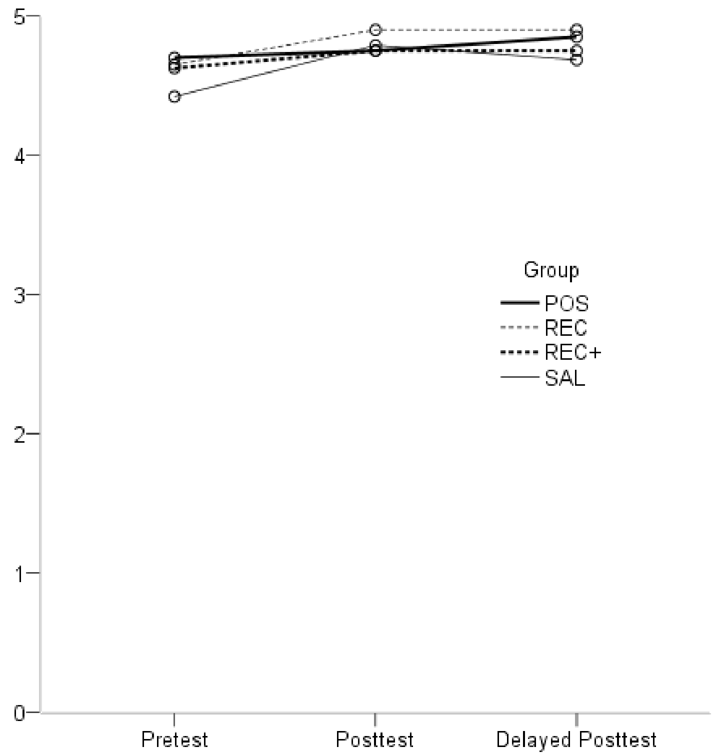


Figure 2. Mean scores of correct picture selection for ungrammatical items (color-size) on the elicited imitation tests. The highest possible score is 5.

Because of a deviation from normality in terms of skewness and kurtosis, nonparametric statistical analyses were performed by using three Kruskal-Wallis tests and four Friedman tests. The alpha level was set at .007 with a Bonferroni adjustment (.05 / 7 tests). First, the Kruskal-Wallis tests showed that the four groups did not differ significantly, $\chi^2(3) = 3.40$, $p = .335$ for the pretest, $\chi^2(3) = 1.90$, $p = .594$ for the posttest, and $\chi^2(3) = 1.22$, $p = .748$ for the delayed posttest. The Friedman tests showed that no group differed across the tests: $\chi^2(2) = 2.00$, $p = .368$ for the POS Group, $\chi^2(2) = 2.25$, $p = .325$ for the REC Group, $\chi^2(2) = 0.29$, $p = .867$ for the REC+ Group, and χ

$^2(2) = 4.77$, $p = .092$ for the SAL Group. In sum, significant differences in the picture selections among the groups were not found.

Imitation analysis. Table 17 shows the descriptive statistics of acceptance of ungrammatical items on the elicited imitation tests (see also Figure 3). A higher score indicates a participant's acceptance of the incorrect order (color-size); a lower score indicates a participant's avoidance of the incorrect order. The mean scores of the REC and REC+ Groups decreased from the pretest to the posttest (2.70 to 2.35 for the REC Group and 3.06 to 2.56 for the REC+ Group). This suggests that both groups decreased the acceptance of the incorrect order (size-color). However, the effects disappeared on the delayed posttest. In contrast, the POS and SAL Groups showed slightly more acceptance of the incorrect order on the posttest (from 3.05 to 3.25 for the POS Group and from 2.79 to 2.95 for the SAL Group) and did not change much between the posttest to the delayed posttest (3.25 to 3.35 for the POS Group and 2.95 to 2.95 for the SAL Group). Thus, the results descriptively indicate that the REC and REC+ treatments resulted in a short-term effect.

Skewness and kurtosis were not significantly large; however, the Kolmogorov-Smirnov tests showed that seven variables (the SAL Group on the pretest, the POS and SAL Groups on the posttest, and the four groups on the delayed posttest) deviated significantly from normality. Thus, I used the same procedure as in the analysis of the oral production tests.

Table 17. *Descriptive Statistics of Incorrect Acceptance of Ungrammatical Items (Color-Size) on the Elicited Imitation Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pretest	3.05	1.15	-0.11	0.51	-0.47	0.99
Posttest	3.25	0.97	-0.17	0.51	0.67	0.99
Delayed	3.35	1.60	-0.56	0.51	-1.31	0.99
REC (*n* = 20)						
Pretest	2.70	1.30	-0.01	0.51	-0.12	0.99
Posttest	2.35	1.66	0.06	0.51	-0.85	0.99
Delayed	3.00	1.72	-0.62	0.51	-0.78	0.99
REC+ (*n* = 16)						
Pretest	3.06	1.39	-0.64	0.56	0.09	1.09
Posttest	2.56	1.26	-0.16	0.56	0.18	1.09
Delayed	3.12	1.54	-0.24	0.56	-0.51	1.09
SAL (*n* = 19)						
Pretest	2.79	1.32	-0.55	0.52	-0.29	1.01
Posttest	2.95	1.35	-0.50	0.52	-0.14	1.01
Delayed	2.95	1.68	-0.53	0.52	-1.08	1.01

Note. The highest possible score is 5. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

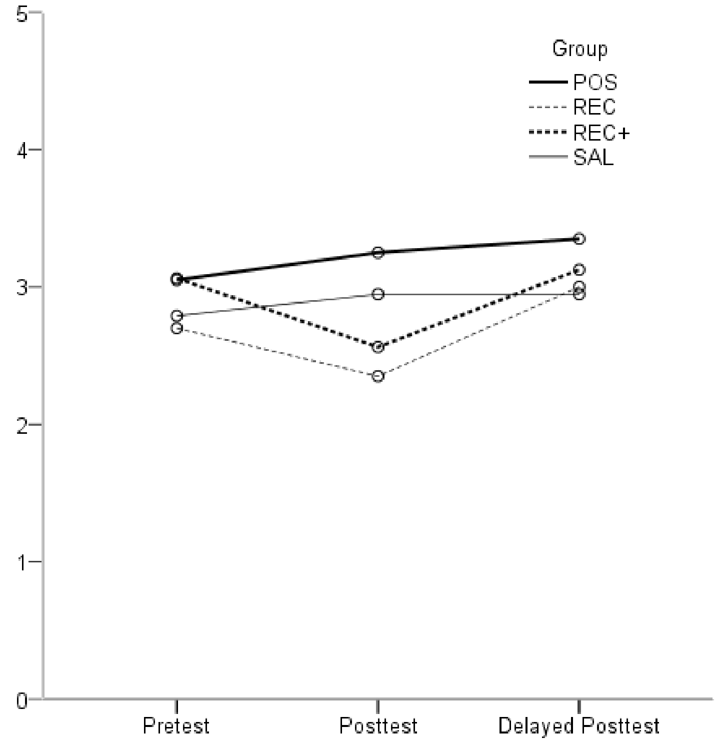


Figure 3. Mean scores of incorrect acceptance of ungrammatical items (color-size) on the elicited imitation tests. The highest possible score is 5.

First, a Friedman test was performed to examine whether the POS Group changed between the tests (3.05, 3.25, and 3.35). No statistically significant differences were found between the tests: $\chi^2(2) = 1.03$, $p = .597$.

Next, difference scores between two of the tests were computed (see Table 18). Because the skewness and kurtosis for one variable (the posttest-delayed posttest period) in the SAL Group deviated from normality, three Kruskal-Wallis tests were performed, instead of parametric tests, to examine whether the four groups differed in their changes between the tests. The alpha level was set at .017 (.05/3 periods) with a Bonferroni adjustment. No statistically significant differences were found among the groups: $\chi^2(3) = 2.35$, $p = .503$ for the pretest-posttest period, $\chi^2(3) = 0.32$, $p = .956$ for the pretest-delayed posttest period, and $\chi^2(2) = 2.87$, $p = .412$ for the posttest-delayed posttest period. Moreover, all the effect sizes for the group comparisons were very small (i.e., less than .10) or small (i.e., less than .30; see Appendix O). Thus, although both the REC and REC+ Groups performed better on the posttest (i.e., the groups decreased their acceptance of the incorrect adjective order from the pretest to the posttest), the statistical analyses did not support this.

Table 18. *Descriptive Statistics of Difference Scores on the Elicited Imitation Tests (Adjective Ordering)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pre-Post	0.20	1.11	-0.18	0.51	-0.75	0.99
Pre-Delayed	0.30	1.49	-0.04	0.51	0.50	0.99
Post-Delayed	0.10	1.59	-0.36	0.51	-1.00	0.99
REC (*n* = 20)						
Pre-Post	-0.35	1.60	-0.22	0.51	-0.69	0.99
Pre-Delayed	0.30	2.15	-0.22	0.51	-0.99	0.99
Post-Delayed	0.65	1.81	0.17	0.51	0.03	0.99
REC+ (*n* = 16)						
Pre-Post	-0.50	1.75	-0.38	0.56	-0.55	1.09
Pre-Delayed	0.06	2.24	0.07	0.56	-0.55	1.09
Post-Delayed	0.56	2.10	-0.37	0.56	1.48	1.09
SAL (*n* = 19)						
Pre-Post	0.16	1.80	-0.45	0.52	0.25	1.01
Pre-Delayed	0.16	1.86	-0.14	0.52	-0.85	1.01
Post-Delayed	0.00	1.41	1.19	0.52	2.48	1.01

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Individual results. For an analysis of individual results, I first examined the individual performances on the pretest of the elicited imitation test and identified how many participants accepted both the correct adjective order and the incorrect order. For this analysis, a participant who accepted 60% or more of the items for both grammatical items (i.e., size-color) and ungrammatical items (i.e., color-size) was considered to have overgeneralized the L1 rule (i.e., Japanese) to the L2 rule (i.e., English). The numbers of items varied across the participants according to the number of correct picture selections; therefore, the cut-off point (60%, which was equivalent to three of the five items) was arbitrarily determined for the judgment of systematicity. In addition, a participant who accepted 40% or less of the items (equivalent to two or fewer of the five items) for ungrammatical items was considered to have shown systematic avoidance of the incorrect adjective order.

Table 19 shows the participants who accepted both the correct order (size-color) and the incorrect order (color-size) on the pretest. The numbers of such participants were 13 out of 20 for the POS Group, 11 out of 20 for the REC Group, 12 out of 16 for the REC+ Group, and nine out of 19 for the SAL Group. Fisher's exact test shows no statistical significance (p = .384) in the frequencies of the participants who had accepted both adjective orders (see Table A in Appendix P).

As shown in Table 19, the performance of the eight participants (1007, 1034, and 1070 of the REC Group, 1017, 1112, and 1114 of the REC+ Group, and 1006 and 1081 of the SAL Group) improved on the posttest, and they maintained their improvement on the delayed test (see the label Retained in Table 19). It should be noted that no one in the POS Group improved on the posttest and maintained the improvement on the delayed posttest whereas there were a few participants who showed maintained improvements (three out of 11 in the REC Groups, three out of 12 in the REC+ Group, and two out of nine in the SAL Group).

Fisher's exact test on the contingency table (Table B in Appendix P) did not show a significant difference in the frequencies of the participants who improved on the posttest and retained that improvement on the delayed test ($p = .188$). Thus, although the numbers of participants who learned ungrammaticality of the color-size order in the REC, REC+ and SAL Groups were larger than that of the POS Group, the difference was not significant.

Even taking delayed effects into consideration, differences between the groups were not clear. Three participants (1013, 1028, and 1092) from the POS Group and one (1064) from the REC+ Group showed delayed effects. According to Fisher's exact test (see Table C in Appendix P), the difference among the groups was not statistically significant ($p = .937$).

Table 19. *Individual Analysis of Performance on the Elicited Imitation Tests (Adjective Ordering)*

		Pretest				Posttest				Delayed posttest			
Participant	Effects	U	%	G	%	U	%	G	%	U	%	G	%
POS (*n* = 20)													
1013	Delayed	3/4	75	4/5	80	3/5	60	4/4	100	1/5	20	5/5	100
1028	Delayed	4/5	80	5/5	100	3/5	60	5/5	100	1/5	20	4/5	80
1092	Delayed	2/3	67	3/5	60	4/4	100	5/5	100	1/5	20	3/4	75
1023	Short	1/1	100	2/3	67	1/3	33	4/4	100	1/2	50	3/3	100
1043		4/5	80	3/4	75	3/5	60	5/5	100	4/5	80	3/4	75
1045		3/5	60	3/3	100	5/5	100	4/5	80	4/4	100	5/5	100
1046		3/5	60	5/5	100	4/5	80	5/5	100	4/5	80	3/4	75
1066		4/5	80	2/3	67	4/5	80	3/5	60	4/4	100	4/5	80
1076		5/5	100	4/4	100	3/4	75	4/5	80	5/5	100	4/5	80
1078		3/3	100	3/5	60	2/4	50	3/4	75	3/5	60	4/5	80
1098		4/4	100	4/5	80	5/5	100	4/5	80	5/5	100	4/5	80
1103		3/5	60	3/3	100	3/5	60	3/5	60	5/5	100	4/5	80
1111		5/5	100	4/5	80	4/5	80	4/4	100	5/5	100	4/4	100
REC (*n* = 20)													
1007	Retained	3/5	60	4/5	80	2/5	40	4/5	80	1/4	25	4/4	100
1034	Retained	4/4	100	3/3	100	2/5	40	2/3	67	2/5	40	3/4	75
1070	Retained	3/5	60	4/5	80	0/5	0	4/4	100	0/5	0	5/5	100
1026	Short	3/5	60	5/5	100	0/5	0	5/5	100	4/5	80	4/5	80
1002		2/2	100	3/4	75	3/4	75	2/5	40	3/5	60	3/5	60
1011		5/5	100	4/5	80	5/5	100	5/5	100	5/5	100	5/5	100
1018		3/4	75	4/5	80	3/5	60	3/5	60	4/5	80	3/5	60
1022		4/5	80	4/4	100	3/4	75	4/5	80	5/5	100	5/5	100
1035		4/4	100	3/5	60	3/4	75	4/5	80	4/5	80	4/5	80
1088		3/4	75	3/4	75	5/5	100	5/5	100	5/5	100	4/5	80
1118		2/2	100	4/4	100	4/5	80	5/5	100	3/5	60	4/5	80
REC+ (*n* = 16)													
1017	Retained	3/5	60	4/4	100	2/5	40	3/5	60	2/5	40	4/5	80
1112	Retained	3/4	75	3/4	75	1/5	20	5/5	100	2/5	40	5/5	100
1114	Retained	5/5	100	3/4	75	2/5	40	4/4	100	2/5	40	4/5	80
1064	Delayed	5/5	100	3/5	60	5/5	100	2/5	40	1/4	25	4/5	80
1019	Short	4/5	80	4/4	100	2/5	40	4/4	100	3/5	60	4/5	80
1115	Short	3/4	75	2/3	67	1/5	20	5/5	100	3/5	60	5/5	100
1021		4/5	80	3/5	60	4/5	80	4/4	100	3/5	60	3/5	60
1051		4/5	80	5/5	100	3/5	60	5/5	100	3/4	75	5/5	100
1072		3/5	60	3/4	75	3/5	60	4/5	80	5/5	100	2/5	40
1087		1/1	100	3/4	75	3/4	75	4/5	80	5/5	100	2/5	40
1090		2/3	67	3/4	75	3/5	60	4/5	80	3/4	75	2/5	40
1116		2/2	100	2/3	67	3/4	75	3/4	75	5/5	100	3/3	100

(*Table 19 continues*)

(*Table 19 continued*)

Participant	Effects	Pretest U	%	G	%	Posttest U	%	G	%	Delayed posttest U	%	G	%
SAL (*n* = 19)													
1006	Retained	4/4	100	5/5	100	2/5	40	3/5	60	2/5	40	4/5	80
1081	Retained	3/3	100	2/3	67	1/5	20	5/5	100	0/5	0	5/5	100
1024		3/3	100	3/3	100	5/5	100	4/5	80	5/5	100	3/4	75
1040		3/5	60	4/4	100	4/5	80	5/5	100	3/5	60	5/5	100
1048		3/3	100	3/5	60	3/4	75	4/5	80	4/4	100	3/4	75
1068		3/5	60	4/5	80	4/5	80	4/4	100	5/5	100	1/5	20
1074		4/4	100	5/5	100	4/5	80	5/5	100	4/5	80	5/5	100
1089		4/5	80	3/5	60	0/0	---	0/0	---	4/5	80	4/5	80
1107		3/4	75	4/5	80	5/5	100	5/5	100	5/5	100	3/4	75

Note. U = ungrammatical items; G = grammatical items. The numerator indicates the number of acceptance; the denominator indicates the number of correct picture selections. Shadowed cells indicate the correct performance, that is, the avoidance of the incorrect order and the acceptance the correct order. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the column of Effects refers to the participant who performed better on the posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Grammaticality Judgment Tests

Group results. The participants were required to judge grammaticality with a four-point rating scale (1 = *Incorrect*; 2 = *Somewhat incorrect*; 3 = *Somewhat correct*; and 4 = *Correct*) for the five ungrammatical items containing the color-size order. The score for each participant was obtained by adding the points for five items and dividing the sum by the number of items. Thus, the score shows each participant's tendency to accept the incorrect color-size-noun sequence. A higher score indicates a participant's acceptance of the incorrect order; a lower score indicates a participant's avoidance of the incorrect order.

Table 20 shows the results for the ungrammatical items on the grammatical judgment tests (see also Figure 4). In general, all the mean scores were above 2.5, suggesting that the participants tended to accept the incorrect order (color-size) across the tests. The POS and SAL Groups did not show a decrease across the tests whereas the REC and REC+ Groups decreased from the pretest to the delayed posttest. The REC Group showed a decline from the pretest to the posttest (3.21 to 2.73) and increased to 3.03 on the delayed posttest. The REC+ Group decreased from the pretest through the posttest (3.51 to 3.19) to the delayed posttest (3.16).

According to Table 20, several of the variables showed significant deviation from normality in terms of skewness and kurtosis. Thus, I used the same procedure as in the analysis of the oral production tests and the elicited imitation tests.

Table 20. *Descriptive Statistics of Incorrect Acceptance of Ungrammatical Items (Color-Size) on the Grammaticality Judgment Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pretest	3.23	0.58	-0.42	0.51	-1.13	0.99
Posttest	3.30	0.81	-1.49	0.51	2.28	0.99
Delayed	3.38	0.86	-1.47	0.51	1.77	0.99
REC (*n* = 20)						
Pretest	3.21	0.56	-0.55	0.51	-0.23	0.99
Posttest	2.73	0.87	-0.13	0.51	-0.93	0.99
Delayed	3.03	0.78	-0.38	0.51	-0.88	0.99
REC+ (*n* = 16)						
Pretest	3.51	0.39	-0.36	0.56	-1.00	1.09
Posttest	3.19	0.84	-1.12	0.56	0.60	1.09
Delayed	3.16	0.69	-0.45	0.56	-0.91	1.09
SAL (*n* = 19)						
Pretest	3.28	0.70	-1.90	0.52	5.74	1.01
Posttest	3.20	0.87	-1.02	0.52	0.58	1.01
Delayed	3.46	0.79	-1.94	0.52	4.30	1.01

Note. The highest possible score is 4; the lowest possible score is 1. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

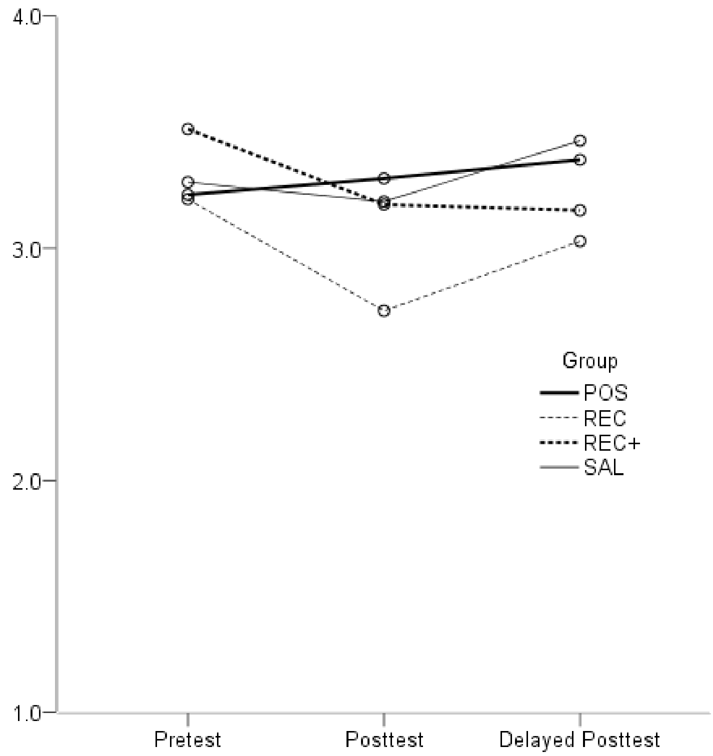


Figure 4. Mean scores of incorrect acceptance of ungrammatical items (color-size) on the grammaticality judgment tests. The possible score range is from 1 to 4.

First, a Friedman test was performed on the scores of the POS Group (3.23 on the pretest, 3.30 on the posttest, and 3.38 on the delayed posttest). There were no statistically significant differences across the tests: $\chi^2(2) = 2.39$, $p = .302$.

Next, the difference scores between two of the three tests were calculated. Table 21 shows the descriptive statistics of the difference scores on the grammaticality judgment test. Because several variables deviated from normality in terms of skewness and kurtosis, I performed nonparametric statistical tests, that is, three Kruskal-Wallis tests. The tests showed no significant differences among the groups for the pretest-posttest period and the posttest-delayed posttest period: $\chi^2(3) = 4.16$, $p = .245$ for the pretest-posttest period and $\chi^2(3) = 2.71$, $p = .439$ for the posttest-delayed posttest period.

Table 21. *Descriptive Statistics of Difference Scores on the Grammaticality Judgment Tests (Adjective Ordering)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 20)						
Pre-Post	0.07	0.70	-0.14	0.51	1.10	0.99
Pre-Delayed	0.15	0.75	-0.91	0.51	2.12	0.99
Post-Delayed	0.08	0.61	-0.08	0.51	-0.15	0.99
REC (*n* = 20)						
Pre-Post	-0.48	0.76	-0.51	0.51	0.58	0.99
Pre-Delayed	-0.18	0.50	-0.32	0.51	-0.59	0.99
Post-Delayed	0.30	0.77	0.05	0.51	0.05	0.99
REC+ (*n* = 16)						
Pre-Post	-0.33	0.89	-1.15	0.56	1.13	1.09
Pre-Delayed	-0.35	0.68	-0.01	0.56	-1.22	1.09
Post-Delayed	-0.03	0.79	0.25	0.56	1.19	1.09
SAL (*n* = 19)						
Pre-Post	-0.08	0.63	-0.20	0.52	0.33	1.01
Pre-Delayed	0.18	0.41	-0.70	0.52	2.17	1.01
Post-Delayed	0.26	0.56	1.18	0.52	2.66	1.01

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

The results for the pretest-delayed posttest period were: $\chi^2(3) = 9.00$, $p = .029$. The difference was not significant because a Bonferroni adjustment was made to set an alpha level at .017 (.05/3 periods). Five comparisons between the groups yielded medium effect sizes (see Appendix O): the comparison between the POS Group and the REC Group ($r = .32$) for the pretest-posttest period and the comparisons (a) between the POS Group and the REC Group ($r = .33$), (b) between the POS Group and the REC+

Group (r = .32), (c) between the REC Group and the SAL Group (r = .37), and (d) between the REC+ Group and the SAL Group (r = .36) for the pretest-delayed posttest period. Therefore, the medium effect sizes suggest that the REC Group improved more than the POS Group for the pretest-posttest period and that the REC and REC+ Groups performed better than the POS and SAL Groups for the pretest-delayed posttest period although the statistical analyses did not yield significance.

Individual results. First, I determined how many participants accepted both the correct adjective order and the incorrect order. As in the analysis of the individual results on the elicited imitation tests, I considered a participant who judged as correct or somewhat correct three or more (i.e., 60%) of the five items for both grammatical items (i.e., size-color) and ungrammatical items (i.e., color-size) to have overgeneralized the Japanese rule to the English rule.

Table 22 shows the participants who accepted both the correct order (size-color) and the incorrect order (color-size) as correct or somewhat correct on the pretest. The numbers of such participants were 15 out of 20 for the POS Group, 13 out of 20 for the REC Group, 16 out of 16 for the REC+ Group, and 15 out of 19 for the SAL Group. Fisher's exact test showed a statistically significant difference among the groups (p = .050; see Table A in Appendix P for the contingency table). In other words, fewer participants in the REC Group than in the other groups accepted both incorrect and correct orders.

According to Table 22, participant 1112 in the REC+ Group and participant 1005 of the SAL Group improved on the posttest (i.e., they rejected three or more of the five ungrammatical items and at the same time accepted three or more of the five grammatical items) and maintained the improvement on the delayed test. The numbers of the participants who improved on the two posttests were quite small: 0 out of 15 and 13 for the POS and REC Groups respectively, 1 out of 16 for the REC+ Group, and 1 out of 15 for the SAL Group. Fisher's exact test showed no statistical significance in the frequencies of the participants who improved on the posttest and maintained the improvement on the delayed test (p = .563; see Table B in Appendix P for the contingency table being analyzed). Thus, no significant differences were observed among the groups.

Taking the delayed effects (i.e., improvements only on the delayed posttest) into consideration (see the label Delayed in Table 22), the numbers of participants showing improvement on the delayed posttest were one (1078) for the POS Group, two (1007 and 1100) for the REC Group, and two (1064 and 1119) for the REC+ Group. Although it seems that the REC and REC+ Groups performed better than the other two groups, Fisher's exact test did not show a statistically significant difference (p = .561; see Table C in Appendix P for the contingency table).

Table 22. *Individual Analysis of Performance on the Grammaticality Judgment Tests (Adjective Ordering)*

		Pretest		Posttest		Delayed posttest	
Participant	Effects	UNGRAM	GRAM	UNGRAM	GRAM	UNGRAM	GRAM
POS (*n* = 20)							
1078	Delayed	4	3	3	2	0	4
1023	Short	5	4	2	3	5	4
1009		5	5	5	5	5	5
1043		5	4	5	5	5	5
1045		5	5	5	5	5	5
1050		3	5	4	4	5	2
1066		5	4	4	4	5	5
1071		5	5	5	5	5	5
1076		4	4	4	5	4	5
1086		5	4	5	5	5	4
1092		5	4	5	5	5	5
1096		4	4	5	5	4	5
1098		5	5	5	5	4	4
1103		3	5	4	2	3	4
1111		4	4	3	4	5	4
REC (*n* = 20)							
1007	Delayed	4	5	3	5	0	5
1100	Delayed	4	3	3	3	2	3
1026	Short	4	4	0	4	3	3
1034	Short	5	4	1	4	5	5
1011		5	4	5	5	5	5
1018		5	5	5	5	5	5
1035		5	5	5	5	5	4
1083		3	5	4	5	4	5
1088		4	4	5	4	5	5
1093		4	3	5	5	5	5
1095		5	4	3	5	5	5
1099		4	5	5	4	4	5
1118		5	3	3	4	4	5
REC (+ *n* = 16)							
1112	Retained	4	4	0	5	0	5
1064	Delayed	3	4	5	2	0	5
1119	Delayed	4	3	4	4	2	4
1019	Short	5	5	0	5	3	5
1069	Short	5	4	0	5	5	0
1012		5	3	4	5	5	3
1017		5	5	5	5	4	5
1021		5	5	5	4	5	5
1051		5	5	5	5	4	5

(*Table 22 continues*)

(*Table 22 continued*)

Participant	Effects	Pretest		Posttest		Delayed posttest	
		UNGRAM	GRAM	UNGRAM	GRAM	UNGRAM	GRAM
1072		3	5	5	5	5	5
1087		5	5	4	5	5	3
1090		5	4	4	4	4	4
1114		5	5	5	5	5	5
1115		4	4	5	5	5	5
1116		5	5	5	5	5	5
1121		4	3	4	5	5	3
SAL (*n* = 19)							
1005	Retained	4	5	0	5	2	3
1006	Short	4	4	1	5	5	0
1010		3	4	4	3	4	4
1014		5	3	4	4	5	5
1040		5	5	5	5	5	5
1048		4	4	4	3	5	1
1068		5	5	5	4	5	5
1074		5	4	5	4	5	5
1080		4	4	3	5	5	1
1081		5	5	5	5	5	5
1089		5	5	5	5	5	5
1102		4	4	3	2	3	2
1107		5	5	4	5	5	5
1108		5	4	5	5	5	5
1110		4	4	5	5	5	5

Note. Shadowed cells indicate the correct performance, that is, the avoidance of the incorrect order and the acceptance the correct order. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the column of Effects refers to the participant who performed better on the posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Summary

In this section, I summarize the results for adjective ordering. The descriptive statistics on the three measures (i.e., visual inspection of the results shown in Figures 1, 3, and 4) suggested that on the elicited imitation tests and grammaticality judgment tests, the REC and REC+ Groups showed short-term effects that paralleled each other; these results differed from those of the POS and SAL Groups. On the other hand, the POS and SAL Groups did not improve, and they changed in a similar way. The REC Group changed differently from the other three groups on the oral production tests.

However, these differences among the groups were not statistically significant. Table 23 summarizes the results of the statistical analyses on the three tests (oral production tests, elicited imitation tests, and grammaticality judgment tests). It must be noted, however, that several comparisons of the groups obtained medium effect sizes on the grammaticality judgment tests. First, the comparison of the POS and REC Groups obtained a medium effect size for the pretest-posttest period. Second, the effect sizes for the comparisons between the two groups (the REC and REC+ Groups) and the other two groups (the POS and the SAL Groups) on the grammaticality judgment tests were medium for the pretest-delayed posttest period. In other words, these medium effect sizes suggest that on the grammaticality judgment tests, the REC Group improved (i.e., rejected ungrammatical adjective ordering) more than the POS Group from the pretest to the posttest and that the REC and REC+ Groups improved from the pretest to the delayed posttest, whereas the POS and SAL Groups did not improve. However, statistical analyses did not show significance.

Table 23. S*ummary of Statistical Analyses for Adjective Ordering*

	Oral production test	Elicited imitation test	Grammaticality judgment test
Friedman tests	*p* = .755	*p* = .597	*p* = .302
Kruskal-Wallis tests			
Pre-Post	*p* = .237	*p* = .503	*p* = .245 Medium ES: REC vs. POS
Pre-Delayed	*p* = .254	*p* = .956	*p* = .029 Medium ES: REC vs. POS REC vs. SAL REC+ vs. POS REC+ vs. SAL
Post-Delayed	*p* = .418	*p* = .412	*p* = .439

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; ES = effect size; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Table 24 summarizes the results of the individual analyses for adjective ordering. The table shows the performance across the three tests by the participants who showed (a) improvement on the posttest and maintenance of that improvement on the delayed posttest (see the label Retained), (b) delayed effects, that is, improvement only on the delayed posttest (see the label Delayed), and (c) short-term effects, that is, improvement only on the posttest (see the label Short). The table also shows information about the pretest for the cells for which the participant did not show improvement. For example, for the POS Group, two (participants 1028 and 1013) of the three participants who improved on the delayed posttest might have possessed correct grammar on the adjective ordering from the start on the basis of their performance on the grammaticality judgment test: Both of them accepted two of the five incorrect items (less than 60% of the items) and accepted all the correct items marked as C in Table 24. One (1013) of them even performed well on the oral production pretest.

Table 24. *Summary of Individual Analyses for Adjective Ordering*

Participant	OPT	EIT	GJT
POS (*n* = 20)			
1078	N (0/0)[a]	B 100/60[b]	Delayed
1028	Delayed	Delayed	C 2/5[c]
1013	C (0/6)	Delayed	C 2/5
1092	N (0/0)	Delayed	B 5/4
1023	N (0/0)	Short	Short
1096	Short	I (50/67)	B 4/4
REC (*n* = 20)			
1034	N (0/0)	Retained	Short
1007	N (0/0)	Retained	Delayed
1070	N (0/0)	Retained	C (2/3)
1083	Retained	I (100/40)	B (3/5)
1100	N (0/0)	I (---/67)	Delayed
1093	Delayed	I (50/100)	B (4/3)
1026	N (0/0)	Short	Short
REC+ (*n* = 16)			
1112	N (0/0)	Retained	Retained
1017	N (0/0)	Retained	B (5/5)
1114	N (0/0)	Retained	B (5/5)
1064	N (0/0)	Delayed	Delayed
1119	N (0/0)	I (100/25)	Delayed
1019	I (3/0)	Short	Short
1069	N (0/0)	I (80/50)	Short
1115	N (0/0)	Short	B (4/4)
SAL (*n* = 19)			
1005	N (0/0)	I (100/50)	Retained
1006	I (1/0)	Retained	Short
1081	I (1/0)	Retained	B (5/5)
1107	Delayed	B (75/80)	B (5/5)
1074	Short	B (100/100)	B (5/4)

Note. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the column of Effects refers to the participant who performed better on the posttest. N = not produced. B = acceptance of both incorrect and correct orders. C = tendency to accept the correct adjective order and avoid the incorrect order, that is, correct judgment or production. I = tendency to accept the incorrect adjective order and avoid the correct order or tendency to accept neither of the orders, that is, incorrect judgment or production. A shadowed cell indicates the improvement of either or both of the posttest and delayed posttest (i.e., retained effects, delayed effects, and short-term effects). POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

[a]The numerator indicates the production of the incorrect adjective order whereas the denominator indicates the frequencies of the correct adjective order.

[b]The numerator indicates the percentage of acceptance of the incorrect adjective order; the denominator indicates the percentage of acceptance of the correct adjective order.

[c]The numerator indicates the number of acceptance of the incorrect adjective order whereas the denominator indicates the number of acceptance of the correct adjective order.

As shown in the previous sections, the four groups did not differ significantly on the three tests on all the measures in terms of the number of participants who produced or accepted both correct and incorrect orders. Table 24 shows that a total of 26 out of the 75 participants (34.7%; six out of 20 in the POS Group, seven out of 20 in the REC Group, eight out of 16 in the REC+ Group, and five out of 19 in the SAL Group) possessed overgeneralized rules on at least one of the three pretests. Moreover, statistically significant differences were not found among the four groups in terms of the number of the participants who improved on the posttest and maintained the improvement on the delayed posttest although it appeared that the numbers of the label Retained in the REC, REC+, and SAL Groups were larger than that in the POS Group.

A comparison of the three tests indicates that the effects seem to be larger on the elicited imitation tests than on the grammaticality judgment tests and that the effects on the oral production tests seem to be the smallest. For example, eight participants were coded as performing better on the posttest and maintaining their improvement on the delayed test (see the label Retained in Table 24) on the elicited imitation tests whereas only one participant and two participants did so on the oral production tests and the grammaticality judgment tests respectively.

The results for adjective ordering are summarized as follows. The first finding concerns the first research question about the effects of exposure to no-contingent positive evidence; the other four findings are related to the other four research questions about the effects of recasts.

1. The POS Group did not perform statistically significantly better or worse on the posttest or the delayed posttest.
2. Both the group analysis and individual analysis showed no statistically significant differences among the four groups on the three measures (oral production tests, elicited imitation tests, and grammaticality judgment tests).
3. Visual inspection of the descriptive statistics showed the following patterns. On the oral production tests, the REC Group improved from the posttest to the delayed posttest whereas the other groups did not change or performed worse. On the elicited imitation tests, the REC and REC+ Groups showed a short-term improvement although the effects disappeared on the delayed posttest. On the grammaticality judgment tests, the REC and REC+ Groups improved on the posttest and maintained the improvements on the delayed test.
4. The individual analyses of the three measures indicated that one third of the participants produced or accepted both the incorrect color-size adjective order and the correct size-color order.
5. Medium effect sizes were observed on the grammaticality judgment tests. More specifically, the improvement of the REC Group from the pretest to the posttest was larger than that of the POS Group, and the improvements of the REC and REC+ Groups were larger than those of the POS and SAL Groups.

Indirect Passives

In the following sections, the results of the three tests for indirect passives are reported.

Oral Production Tests

Group results. Table 25 shows the descriptive statistics regarding the frequencies of indirect passives, an ungrammatical form in English, for the oral production tests (see also Figure 5). As shown in Table 25, the assumption of normality in the distribution of each level of each test was not observed in terms of skewness and kurtosis. Therefore, a nonparametric Friedman test was performed to examine whether the POS Group changed between the tests. Following this, statistical tests were performed to examine whether the difference scores between the tests differed between groups.

The participants in the POS Group produced 0.14 instances of indirect passives on average on the pretest. The mean score increased slightly to 1.33 on the posttest and decreased slightly to 1.10 on the delayed posttest (This trend was observed for the other groups as well; see Figure 5). The Friedman test showed significant differences between the tests, $\chi^2(2) = 14.00$, $p = .001$. To compare two of the tests, I performed three Wilcoxon signed-rank tests. The tests showed significant differences between the pretest and the posttest ($z = -2.69$, $p = .007$) and between the pretest and the delayed posttest ($z = -2.39$, $p = .017$). There was no significant difference between the posttest and the delayed posttest ($z = -1.67$, $p = .096$).

Table 25. *Descriptive Statistics of the Frequencies of Indirect Passives on the Oral Production Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pretest	0.14	0.36	2.20	0.50	3.14	0.97
Posttest	1.33	1.71	0.68	0.50	-1.44	0.97
Delayed	1.10	1.67	1.04	0.50	-0.74	0.97
REC (*n* = 22)						
Pretest	0.27	0.88	3.06	0.49	8.09	0.95
Posttest	1.36	1.59	0.59	0.49	-1.38	0.95
Delayed	0.82	1.50	1.55	0.49	0.76	0.95
REC+ (*n* = 27)						
Pretest	0.33	0.62	1.74	0.45	2.08	0.87
Posttest	2.00	1.64	-0.28	0.45	-1.71	0.87
Delayed	0.78	1.40	1.62	0.45	1.20	0.87
SAL (*n* = 20)						
Pretest	0.40	1.05	2.74	0.51	7.40	0.99
Posttest	1.20	1.64	0.99	0.51	-0.76	0.99
Delayed	0.95	1.70	1.30	0.51	-0.29	0.99

Note. The highest possible score is 4. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

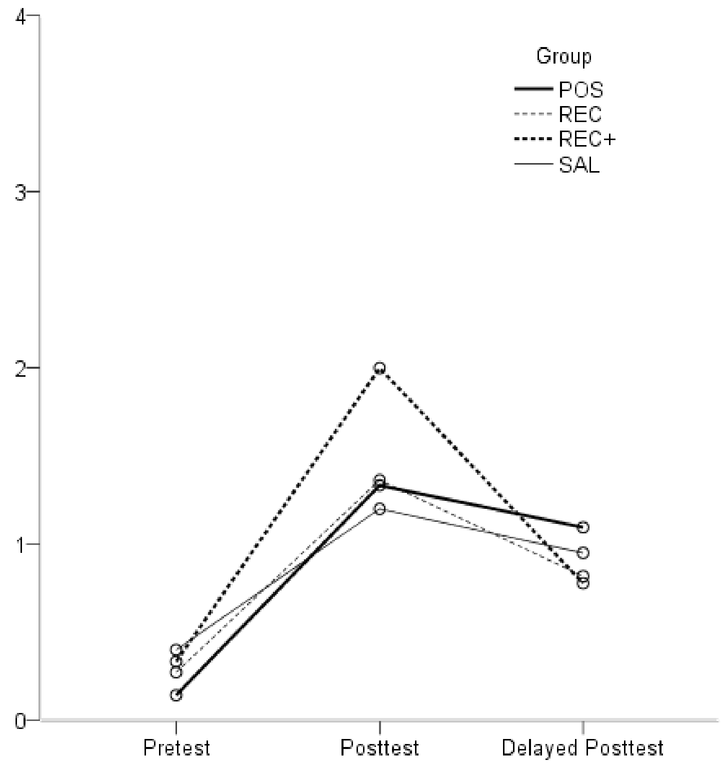


Figure 5. Mean frequencies of indirect passives on the oral production tests. The highest possible score is 4.

Table 26 presents the descriptive statistics for the difference scores between two of the three oral production tests. The distribution of each variable deviated from normality in terms of skewness and kurtosis. Therefore, three Kruskal-Wallis tests were performed to compare the groups on the three tests. The alpha level was set at .017 (.05/3 periods) with a Bonferroni adjustment. The tests showed no statistically significant differences between the groups: $\chi^2(3) = 3.62$, $p = .306$ for the pretest-posttest period; $\chi^2(3) = 1.93$, $p = .586$ for the pretest-delayed posttest period; and $\chi^2(3) = 4.47$, $p = .215$ for the posttest-delayed posttest period. Furthermore, effect sizes for the comparisons between the groups were very small (i.e., less than .10) or small (i.e., less than .30; see Appendix Q). Thus, Table 26 (see also Figure 5) shows that the REC+ Group decreased the use of indirect passives for the posttest-delayed posttest period (mean difference score = -1.22) to a greater extent than the other groups (-0.24, -0.55, and -0.25 for the POS, REC, and SAL Groups, respectively); however, the differences were not statistically significant.

Table 26. *Descriptive Statistics of Difference Scores on the Oral Production Tests (Indirect Passives)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pre-Post	1.19	1.54	0.74	0.50	-1.18	0.97
Pre-Delayed	0.95	1.47	1.14	0.50	-0.25	0.97
Post-Delayed	-0.24	0.62	-1.16	0.50	2.42	0.97
REC (*n* = 22)						
Pre-Post	1.09	1.72	0.53	0.49	-0.82	0.95
Pre-Delayed	0.55	1.79	0.71	0.49	0.63	0.95
Post-Delayed	-0.55	1.10	-1.29	0.49	0.54	0.95
REC+ (*n* = 27)						
Pre-Post	1.67	1.57	0.22	0.45	-1.63	0.87
Pre-Delayed	0.44	1.37	1.05	0.45	0.94	0.87
Post-Delayed	-1.22	1.78	-0.25	0.45	-1.17	0.87
SAL (*n* = 20)						
Pre-Post	0.80	1.58	1.08	0.51	0.94	0.99
Pre-Delayed	0.55	1.57	1.57	0.51	2.01	0.99
Post-Delayed	-0.25	1.41	0.37	0.51	6.24	0.99

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Individual results. The individual performances on the production task pretest were examined to check how many participants produced both correct passives and indirect passives, which might be due to L1 influence. As shown in Table 27, three out of 21 participants in the POS Group, two out of 22 participants in the REC Group, seven out of 27 participants in the REC+ Group, and two out of 20 participants in the SAL Group showed both correct production (correct use of passive or active voice or the *have* + *pp* construction) and incorrect production. Fisher's exact test (see Table A in Appendix R for the contingency table) showed that the distributions were not significantly different among the groups (p = .394).

Of the 14 participants, one (1020) in the SAL Group did not produce the indirect passives on the posttest and continued no-use of the indirect passives on the delayed posttest (see the label Retained in Table 27). Five participants (1035 in the REC Group and 1049, 1064, 1091, and 1119 in the REC+ Group) did not produce the indirect passives on the delayed posttest; these five participants showed delayed effects of the treatments.

First, a Fisher's exact test was performed on the numbers of the participants who did not produce the indirect passives on the posttest or on the delayed posttest (see the label Retained in Table 27; see Table B in Appendix R for the contingency table), and no statistically significant difference among the groups (p = .286) was found. Second, another Fisher's exact test was performed, considering the participants who showed delayed effects (see the label Delayed in Table 27; see Table C in Appendix R for the contingency table). There was no statistically significant difference among the groups (p = .099). Thus, although in the three experimental groups (REC, REC+, and SAL), no-use of the indirect passives either on the two posttests or on the delayed posttest was found for six participants in the REC, REC+, and SAL Groups and the three participants of the POS Group kept producing indirect passives, the statistical test did not support this difference.

Table 27. *Individual Analysis of Performance on the Oral Production Tests (Passivization)*

		Pretest			Posttest			Delayed posttest		
Participant	Effects	COR	IDP	Other	COR	IDP	Other	COR	IDP	Other
POS (*n* = 21)										
1009		3	1	0	0	4	0	1	3	0
1050		3	1	0	0	4	0	0	4	0
1092		2	1	1	1	3	0	0	4	0
REC (*n* = 22)										
1035	Delayed	1	3	0	2	1	1	4	0	0
1042		1	3	0	1	3	0	3	1	0
REC+ (*n* = 27)										
1049	Delayed	2	1	1	1	3	0	4	0	0
1064	Delayed	3	1	0	2	2	0	4	0	0
1091	Delayed	2	1	1	0	4	0	4	0	0
1119	Delayed	2	2	0	1	3	0	4	0	0
1012		3	1	0	1	3	0	3	1	0
1037		2	1	1	2	2	0	0	4	0
1090		1	2	1	1	3	0	0	4	0
SAL (*n* = 20)										
1020	Retained	2	2	0	4	0	0	4	0	0
1068		1	2	1	0	3	1	0	3	1

Note. COR = correct use of passives (i.e., active voice) or have + pp construction; IDP = indirect passives; Other = no use of passives. The shadowed cells indicate no-use of IDP. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Elicited Imitation Tests

Group results. As explained in the Data Coding section, imitation performance was analyzed with the correct selection of a picture. Therefore, I first checked whether the participants' selection of pictures for the ungrammatical items of the elicited imitation tests differed across the groups. Imitation performance was then analyzed.

Picture selection. Table 28 shows the descriptive statistics (see also Figure 6). In general, the scores were high, ranging from 3.85 to 4.76, and the standard deviations were small, ranging from 0.51 to 1.13. The performances on the picture selection suggest that the participants focused on the meanings of the items and comprehended the sentences well.

Table 28. *Descriptive Statistics of Correct Picture Selection for Ungrammatical Items (Indirect Passives) on the Elicited Imitation Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pretest	4.48	0.75	-1.09	0.50	-0.20	0.97
Posttest	4.76	0.54	-2.32	0.50	5.06	0.97
Delayed	4.62	0.50	-0.53	0.50	-1.91	0.97
REC (*n* = 22)						
Pretest	3.86	0.99	-0.68	0.49	-0.32	0.95
Posttest	4.50	0.60	-0.74	0.49	-0.31	0.95
Delayed	4.32	0.78	-0.65	0.49	-0.99	0.95
REC+ (*n* = 27)						
Pretest	3.85	1.13	-0.89	0.45	0.45	0.87
Posttest	4.15	0.66	-0.17	0.45	-0.57	0.87
Delayed	4.22	0.80	-0.92	0.45	0.76	0.87
SAL (*n* = 20)						
Pretest	3.95	0.94	-1.56	0.51	4.11	0.99
Posttest	4.55	0.69	-1.28	0.51	0.54	0.99
Delayed	4.50	0.51	0.00	0.51	-2.24	0.99

Note. The highest possible score is 5. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

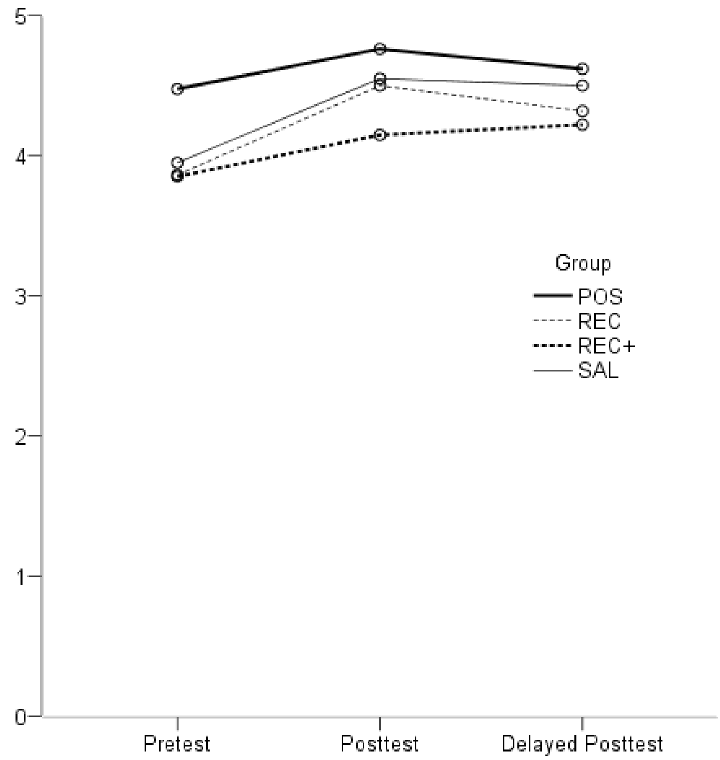


Figure 6. Mean scores of correct picture selection for ungrammatical items (indirect passives) on the elicited imitation tests. The highest possible score is 5.

Because of the deviation from normality in terms of skewness and kurtosis, nonparametric statistical tests (three Kruskal-Wallis tests and four Friedman tests) were performed. The alpha level was set at .007 with a Bonferroni adjustment (.05/7 tests). The Kruskal-Wallis tests showed that the four groups did not differ significantly on the pretest and the delayed posttest ($\chi^2(3) = 6.44$, $p = .092$ for the pretest and $\chi^2(3) = 3.45$, $p = .328$ for the delayed posttest); however, they differed significantly on the posttest ($\chi^2(3) = 12.52$, $p = .006$). A post hoc analysis with Mann-Whitney tests showed a statistically significant difference between the POS Group and the REC+ Group. Friedman tests showed that the POS, REC, and REC+ Groups did not differ across the tests: $\chi^2(2) = 2.45$, $p = .294$ for the POS Group, $\chi^2(2) = 6.62$, $p = .037$ for the REC Group, $\chi^2(2) = 2.67$, and $p = .264$ for the REC+ Group. However, the SAL Group changed significantly between tests: $\chi^2(2) = 14.53$, $p = .001$. A post hoc analysis using Wilcoxon signed-rank tests did not show significant difference for any combination of tests. In sum, statistical differences in the picture selections between the groups were found, as the POS Group performed better than REC+ Group on the posttest. However, these differences were not believed to influence the following imitation analysis because the imitation analysis was based on the imitation performance on the items for which the participants selected correct pictures.

Imitation analysis. Table 29 shows the descriptive statistics for the acceptance of ungrammatical items on the elicited imitation tests (see also Figure 7). A higher score indicates a participant's incorrect acceptance of indirect passives; a lower score indicates a participant's avoidance of indirect passives. All groups accepted indirect passives on the posttest more frequently than on the pretest. In particular, the mean score of the REC Group showed an increase from 2.86 to 4.05. The mean scores of the REC and REC+ Groups showed decreases from the posttest to the delayed posttest (4.05 to 3.77 for the REC Group and 3.26 to 3.19 for the REC+ Group). On the other hand, the POS and SAL Groups' tendency to accept indirect passives became stronger on the delayed posttest (3.62 to 3.86 for the POS Group and 3.55 to 3.85 for the SAL Group).

Table 29. *Descriptive Statistics of Incorrect Acceptance of Ungrammatical Items (Indirect Passives) on the Elicited Imitation Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pretest	3.57	1.40	-0.59	0.50	-1.02	0.97
Posttest	3.62	1.66	-1.21	0.50	0.37	0.97
Delayed	3.86	1.20	-1.83	0.50	4.60	0.97
REC (*n* = 22)						
Pretest	2.86	1.04	-0.27	0.49	0.49	0.95
Posttest	4.05	1.05	-1.20	0.49	1.80	0.95
Delayed	3.77	0.97	-0.18	0.49	-0.94	0.95
REC+ (*n* = 27)						
Pretest	2.74	1.29	-0.64	0.45	-0.55	0.87
Posttest	3.26	1.40	-0.95	0.45	0.39	0.87
Delayed	3.19	1.57	-0.97	0.45	-0.13	0.87
SAL (*n* = 20)						
Pretest	2.70	1.45	-0.44	0.51	-1.54	0.99
Posttest	3.55	1.10	-0.80	0.51	0.10	0.99
Delayed	3.85	1.09	-1.03	0.51	1.09	0.99

Note. The highest possible score is 5. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

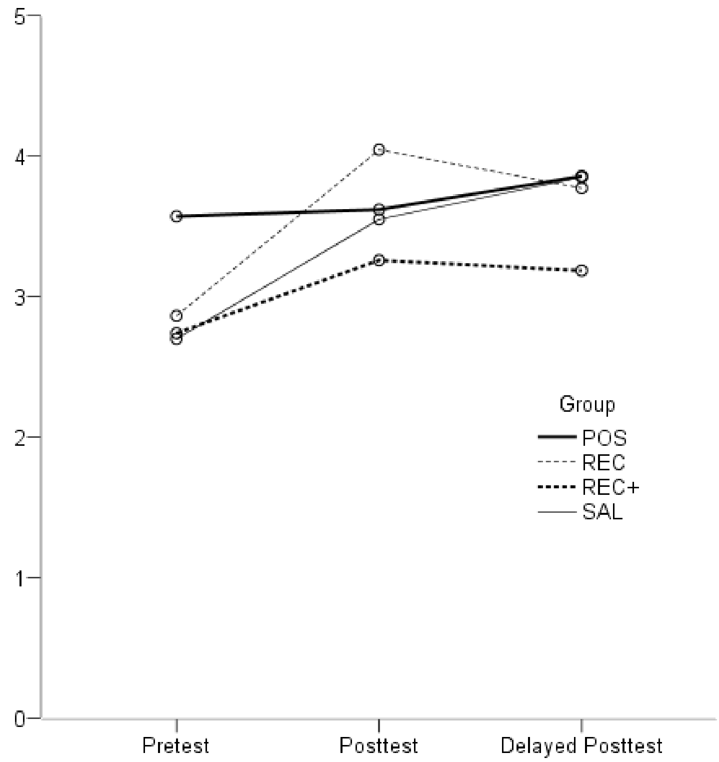


Figure 7. Mean scores of incorrect acceptance of ungrammatical items (indirect passives) on the elicited imitation tests. The highest possible score is 5.

Because several variables deviated from normality in terms of skewness and kurtosis, the same procedure used for the analysis of the oral production tests was applied. First, a Friedman test was performed to examine whether the POS Group differed among the tests. No significant difference was found, $\chi^2(2) = 0.27$, $p = .874$.

Table 30 shows the descriptive statistics of difference scores between two of the three tests. Because some skewness and kurtosis in the difference scores indicated deviations from normality, three Kruskal-Wallis tests were performed, instead of parametric statistical tests, to examine whether the four groups differed in their changes between the tests. The alpha level was set at .017 (.05/3 periods) with a Bonferroni adjustment. No statistically significant differences were found: $\chi^2(3) = 7.95$, $p = .047$ for the pretest-posttest period, $\chi^2(3) = 5.98$, $p = .113$ for the pretest-delayed posttest period, and $\chi^2(2) = 4.40$, $p = .221$ for the posttest-delayed posttest period. Medium effect sizes were observed for the two comparisons (see Appendix Q): between the POS Group and the REC Group ($r = .41$) for the pretest-posttest period and between the POS Group and the SAL Group ($r = .36$) for the pretest-delayed posttest period. Thus, the REC Group increased in their tendency to accept ungrammatical indirect passives from

Table 30. *Descriptive Statistics of Difference Scores on the Elicited Imitation Tests (Indirect Passives)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pre-Post	0.05	1.40	-0.34	0.50	-0.20	0.97
Pre-Delayed	0.29	1.35	0.64	0.50	-0.07	0.97
Post-Delayed	0.24	1.09	0.24	0.50	-0.11	0.97
REC (*n* = 22)						
Pre-Post	1.18	1.01	0.53	0.49	-0.62	0.95
Pre-Delayed	0.91	1.06	0.46	0.49	-0.39	0.95
Post-Delayed	-0.27	0.94	-0.55	0.49	-0.30	0.95
REC+ (*n* = 27)						
Pre-Post	0.52	1.53	-0.00	0.45	0.39	0.87
Pre-Delayed	0.44	1.60	-0.98	0.45	1.64	0.87
Post-Delayed	-0.07	1.07	-1.87	0.45	5.91	0.87
SAL (*n* = 20)						
Pre-Post	0.85	1.09	0.33	0.51	-1.06	0.99
Pre-Delayed	1.15	1.14	0.63	0.51	-0.94	0.99
Post-Delayed	0.30	0.57	-0.04	0.51	0.40	0.99

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

the pretest to the posttest more than the POS Group, and the SAL Group performed worse (i.e., accepted more ungrammatical indirect passives) from the pretest to the delayed posttest, whereas the POS Group did not change across the three testing times.

Individual results. First, individual results of the pretest of the elicited imitation test were examined and the number of participants who accepted both correct passives and incorrect indirect passives was tallied. For this analysis, 60% was set as an arbitrary cut-off point for the judgment of the participant's systematicity as in the individual analysis of the elicited imitation for adjective ordering. Thus, participants who accepted both grammatical items (60% or more of the items) and ungrammatical items (60% or more of the items) were considered to have overgeneralized the L1 rule (i.e., Japanese) to the L2 rule (i.e., English). In addition, participants who accepted 40% or less of the items were considered as showing systematic avoidance of the indirect passives. Table 31 shows the participants who accepted both correct passives and indirect passives on the pretest.

The numbers of such participants were 12 out of 21 in the POS Group, 14 out of 22 in the REC Group, 17 out of 27 in the REC+ Group, and 13 out of 20 in the SAL Group. Fisher's exact test (see Table A in Appendix R for the contingency table) showed that there was no significant difference in the frequencies of participants who had accepted both correct passives and incorrect indirect passives (p = .972).

Table 31. *Individual Analysis of Performance on the Elicited Imitation Tests (Passivization)*

		Pretest				Posttest				Delayed posttest			
Participant	Effects	U	%	G	%	U	%	G	%	U	%	G	%
POS (*n* = 21)													
1043	Short	4/5	80	2/3	67	1/4	25	4/5	80	3/4	75	4/4	100
1009		5/5	100	1/1	100	4/5	80	3/4	75	4/4	100	3/3	100
1028		5/5	100	1/1	100	4/4	100	2/2	100	4/4	100	4/4	100
1029		5/5	100	3/3	100	5/5	100	3/3	100	4/4	100	3/3	100
1045		4/5	80	2/2	100	4/5	80	1/1	100	5/5	100	3/4	75
1050		4/4	100	3/3	100	4/5	80	4/4	100	4/4	100	3/3	100
1066		5/5	100	4/4	100	5/5	100	3/3	100	5/5	100	4/4	100
1071		4/4	100	4/4	100	5/5	100	4/5	80	5/5	100	5/5	100
1076		4/4	100	2/2	100	4/4	100	1/1	100	4/4	100	2/2	100
1098		4/5	80	3/3	100	5/5	100	3/4	75	4/5	80	3/4	75
1103		3/3	100	1/1	100	5/5	100	1/1	100	4/4	100	2/3	67
1109		5/5	100	1/1	100	4/4	100	1/1	100	4/4	100	1/1	100
REC (*n* = 22)													
1002		4/4	100	1/1	100	5/5	100	2/2	100	4/4	100	2/2	100
1004		3/4	75	4/4	100	4/4	100	5/5	100	4/4	100	5/5	100
1007		4/5	80	4/4	100	4/4	100	4/4	100	5/5	100	5/5	100
1011		4/5	80	2/2	100	4/4	100	3/4	75	4/4	100	4/4	100
1026		4/5	80	1/1	100	5/5	100	3/3	100	3/4	75	4/4	100
1035		4/5	80	3/4	75	5/5	100	3/4	75	4/4	100	3/3	100
1038		2/2	100	2/2	100	3/3	100	3/3	100	3/3	100	5/5	100
1042		2/3	67	3/3	100	5/5	100	4/4	100	5/5	100	4/4	100
1065		3/3	100	3/3	100	5/5	100	3/4	75	5/5	100	4/4	100
1073		4/5	80	2/2	100	5/5	100	3/3	100	5/5	100	3/3	100
1083		4/4	100	4/4	100	4/4	100	3/4	75	4/4	100	5/5	100
1088		4/4	100	2/2	100	4/4	100	3/3	100	4/4	100	2/2	100
1117		3/4	75	1/1	100	3/5	60	3/3	100	3/4	75	2/2	100
1118		3/3	100	1/1	100	5/5	100	2/3	67	3/3	100	3/3	100
REC+ (*n* = 27)													
1051	Retained	4/5	80	5/5	100	1/4	25	5/5	100	1/5	20	5/5	100
1003		4/4	100	3/3	100	5/5	100	3/4	75	4/4	100	4/4	100
1012		4/4	100	4/4	100	4/4	100	4/4	100	0/0		0/0	
1017		2/2	100	3/4	75	4/4	100	3/4	75	4/4	100	4/5	80
1030		4/4	100	2/2	100	4/4	100	4/4	100	4/4	100	4/4	100

(*Table 31 continues*)

(*Table 31 continued*)

		Pretest				Posttest				Delayed posttest			
Participant	Effects	U	%	G	%	U	%	G	%	U	%	G	%
1036		4/4	100	2/2	100	4/4	100	4/4	100	5/5	100	4/4	100
1049		3/4	75	1/1	100	4/4	100	2/3	67	4/4	100	4/4	100
1064		4/4	100	4/4	100	5/5	100	3/3	100	5/5	100	3/3	100
1075		3/3	100	5/5	100	2/3	67	3/4	75	2/4	50	4/4	100
1077		4/5	80	2/2	100	3/4	75	3/3	100	4/4	100	2/2	100
1094		4/4	100	4/4	100	4/4	100	3/4	75	4/4	100	4/4	100
1112		3/3	100	4/5	80	4/4	100	2/3	67	5/5	100	4/5	80
1115		4/5	80	1/1	100	5/5	100	1/2	50	4/5	80	2/3	67
1116		2/3	67	2/2	100	4/4	100	1/2	50	3/3	100	2/2	100
1119		4/4	100	2/2	100	3/3	100	0/2	0	4/4	100	2/2	100
1120		2/2	100	2/2	100	2/3	67	3/3	100	3/4	75	5/5	100
1121		2/3	67	1/1	100	4/4	100	2/2	100	4/5	80	2/2	100
SAL (*n* = 20)													
1005		4/4	100	2/2	100	5/5	100	2/3	67	5/5	100	3/4	75
1006		4/4	100	3/3	100	4/4	100	3/3	100	4/4	100	4/4	100
1014		4/4	100	4/4	100	4/4	100	2/2	100	5/5	100	5/5	100
1020		4/5	80	3/3	100	4/5	80	1/1	100	5/5	100	3/4	75
1040		4/5	80	3/3	100	5/5	100	3/4	75	5/5	100	4/4	100
1041		4/4	100	3/3	100	4/4	100	5/5	100	4/4	100	5/5	100
1048		4/4	100	2/2	100	4/4	100	4/4	100	4/4	100	3/3	100
1068		4/5	80	3/3	100	5/5	100	3/4	75	5/5	100	4/4	100
1080		4/4	100	4/4	100	4/4	100	4/4	100	4/4	100	4/4	100
1081		2/3	67	2/2	100	4/4	100	3/3	100	5/5	100	2/3	67
1089		3/4	75	4/4	100	2/4	50	0/0		3/5	60	4/4	100
1108		4/4	100	3/3	100	4/4	100	4/4	100	4/4	100	2/3	67
1110		1/1	100	1/1	100	3/3	100	2/3	67	4/4	100	2/2	100

Note. U = ungrammatical items; G = grammatical items. The numerator indicates the number of acceptance; the denominator indicates the number of correct picture selection. The shadowed cells indicate the correct performance, that is, the avoidance of the indirect passives and the acceptance the correct passives. Retained (i.e., retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Short (i.e., short-term effects) in the column of Effects refers to the participant who performed better on the posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

As shown in Table 31, participant 1051 in the REC+ Group improved on the posttest and maintained the improvement on the delayed test. No participants from the other three groups showed improvement. Fisher's exact test (see Table B in Appendix R for the contingency table) showed no significant difference in the frequencies of the participants who improved on the posttest and whose improvement was maintained on the delayed test (p = 1.000). No participants showed delayed effects.

Grammaticality Judgment Tests

Group results. The participants were asked to judge grammaticality with a four-point rating scale (1 = *Incorrect*; 2 = *Somewhat incorrect*; 3 = *Somewhat correct*; and 4 = *Correct*). There were five ungrammatical items (i.e., indirect passives). The score for each participant was obtained by adding the points for five items and dividing the sum by the number of items. Thus, the score shows each participant's tendency to accept the incorrect color-size-noun sequence. A higher score indicates a participant's acceptance of the incorrect order; a lower score indicates a participant's avoidance of the incorrect order.

Table 32 shows the descriptive statistics for the ungrammatical items on the grammatical judgment tests (see also Figure 8). In general, the mean scores of the four groups increased from the pretest to the posttest. The REC+ and SAL Groups decreased from the posttest to the delayed posttest (3.18 to 2.92 for the REC+ Group and 2.87 to 2.60 for the SAL Group), whereas the POS and REC Groups decreased very little from the posttest to the delayed posttest.

Table 32. *Descriptive Statistics of Incorrect Acceptance of Ungrammatical Items (Indirect Passives) on the Grammaticality Judgment Tests*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pretest	2.45	0.78	0.55	0.50	-0.25	0.97
Posttest	3.01	0.96	-1.09	0.50	0.40	0.97
Delayed	2.97	0.92	-0.87	0.50	0.13	0.97
REC (*n* = 22)						
Pretest	2.65	0.70	0.22	0.49	-0.66	0.95
Posttest	3.36	0.69	-0.59	0.49	-1.22	0.95
Delayed	3.25	0.63	-0.42	0.49	-1.11	0.95
REC+ (*n* = 27)						
Pretest	2.77	0.55	-0.04	0.45	-1.26	0.87
Posttest	3.18	0.82	-1.15	0.45	0.45	0.87
Delayed	2.92	0.67	-0.59	0.45	-0.32	0.87
SAL (*n* = 20)						
Pretest	2.60	0.74	-0.65	0.51	-0.20	0.99
Posttest	2.87	0.86	-0.58	0.51	-0.36	0.99
Delayed	2.60	0.91	-0.17	0.51	-0.91	0.99

Note. The highest possible score is 4; the lowest possible score is 1. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

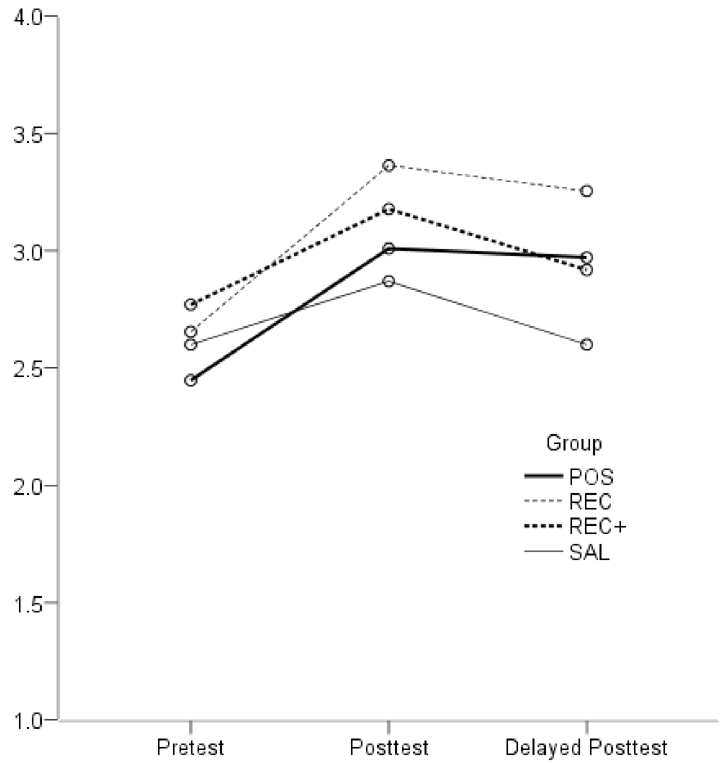


Figure 8. Mean scores of incorrect acceptance of ungrammatical items (indirect passives) on the grammaticality judgment tests. The possible score range is from 1 to 4.

As shown in Table 33, several of the variables showed significant deviation from normality in terms of skewness. Thus, the same analytical procedure used with the oral production tests and the elicited imitation tests was applied in this case.

First, a Friedman test was performed on the scores of the POS Group. Statistically significant differences were found across the tests: $\chi^2(2) = 10.49$, $p = .005$. In order to investigate the differences between two of the three tests, three Wilcoxon signed-rank tests were performed. The pretest and the posttest were significantly different ($z = -2.82$, $p = .005$) as was the pretest and the delayed posttest ($z = -3.05$, $p = .002$). The difference between the posttest and the delayed posttest was not significant: $z = -0.08$, $p = .936$. In other words, the POS Group performed worse on the posttest than on the pretest and remained at the same level on the delayed posttest. This trend was observed for the other three groups as well.

Table 33. *Descriptive Statistics of Difference Scores on the Grammaticality Judgment Tests (Indirect Passives)*

	M	*SD*	Skewness	*SES*	Kurtosis	*SEK*
POS (*n* = 21)						
Pre-Post	0.56	0.76	-0.24	0.50	-1.07	0.97
Pre-Delayed	0.53	0.61	-0.36	0.50	-0.52	0.97
Post-Delayed	-0.04	0.42	-1.55	0.50	4.84	0.97
REC (*n* = 22)						
Pre-Post	0.71	0.76	0.76	0.49	0.58	0.95
Pre-Delayed	0.60	0.70	0.80	0.49	-0.01	0.95
Post-Delayed	-0.11	0.62	-0.98	0.49	1.61	0.95
REC+ (*n* = 27)						
Pre-Post	0.41	0.91	-1.15	0.45	3.64	0.87
Pre-Delayed	0.15	0.73	-0.67	0.45	1.83	0.87
Post-Delayed	-0.26	0.49	-0.82	0.45	0.90	0.87
SAL (*n* = 20)						
Pre-Post	0.27	0.82	-1.13	0.51	1.37	0.99
Pre-Delayed	0.00	0.91	-1.10	0.51	1.25	0.99
Post-Delayed	-0.27	0.63	0.25	0.51	0.42	0.99

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Next, the difference scores between two of the three tests were calculated. Table 33 shows the descriptive statistics for the difference scores on the grammaticality judgment test. Because some of the variables in Table 33 deviated from normality in terms of skewness and kurtosis, three Kruskal-Wallis tests were performed instead of parametric tests. The alpha level was set at .017 (.05/3 periods) with a Bonferroni adjustment. The tests showed no significant differences among the groups for the pretest-posttest period, the pretest-delayed posttest period, and the posttest-delayed posttest period: $\chi^2(3) = 2.10$, $p = .553$ for the pretest-posttest period, $\chi^2(3) = 7.14$, $p = .068$ for the pretest-delayed posttest period, and $\chi^2(3) = 4.18$, $p = .242$ for the posttest-delayed posttest period. Moreover, all the effect sizes for the group comparisons except for one comparison between the POS and SAL Groups for the pretest-delayed posttest period were very small (i.e., less than .10) or small (i.e., less than .30; see Appendix Q). The comparison between the POS and SAL Groups for the pretest-delayed posttest period showed a medium effect size; in other words, it was suggested that the change of the POS Group from the pretest to the delayed posttest was larger than that of the SAL Group.

Individual results. First, the number of participants who accepted both correct passives and incorrect indirect passives on the posttest was examined. Like in the analysis of the individual results for the elicited imitation tests, participants who judged as correct or somewhat correct three or more (i.e., 60% or more) of the five items for both grammatical items and ungrammatical items were considered to have overgeneralized the Japanese rule to the English rule. In addition, participants who accepted two or less (i.e., 40% or less) of the items were considered as showing the systematic avoidance of the indirect passives.

Table 34 shows the participants who accepted both the correct passives and the incorrect indirect passives as correct or somewhat correct on the pretest: five out of 21 participants in the POS Group, eight out of 22 participants in the REC Group, 11 out of 27 participants in the REC+ Group, and seven out of 20 participants in the SAL Group. Fisher's exact test (see Table A in Appendix R for the contingency table) showed that there was no statistical significance in the frequencies of the participants who had accepted both grammatical and ungrammatical passives (p = .679).

As shown in Table 34, three participants (1051 and 1112 in the REC+ Group and 1089 in the SAL Group) improved on the posttest and maintained the improvement on the delayed test (see the label Retained). Fisher's exact test (see Table B in Appendix R) was performed; no significant difference was found in the frequencies of the participants who improved on the posttest and retained the improvement on the delayed test (p = .667). Thus, no differences were observed among the groups. Two of the participants showed improvement on the delayed posttest (see the label Delayed in Table 34), participant 1075 in the REC+ Group and participant 1040 in the SAL Group. Taking the participants who showed delayed effects into consideration, a Fisher's exact test was performed (see Table C in Appendix R); no significant difference was found (p =. 292).

Table 34. *Individual Analysis of Performance on the Grammaticality Judgment Tests (Passivization)*

		Pretest		Posttest		Delayed Posttest	
Participant	Effects	UNGRAM	GRAM	UNGRAM	GRAM	UNGRAM	GRAM
POS (*n* = 21)							
1009		5	3	5	4	5	4
1013		4	3	5	4	5	4
1092		3	4	5	3	5	4
1109		3	3	5	5	5	5
1028		3	3	2	2	3	3
REC (*n* = 22)							
1004		3	3	5	2	5	4
1007		5	4	5	5	5	5
1011		3	3	5	5	5	4
1026		3	4	5	4	4	3
1035		4	4	4	3	4	3
1042		4	3	5	4	3	3
1065		4	4	4	4	5	5
1085		3	3	5	4	5	3
REC+ (*n* = 27)							
1051	Retained	4	5	0	3	1	3
1112	Retained	3	5	0	5	0	4
1075	Delayed	3	5	5	4	1	5
1017		4	5	5	5	5	5
1030		4	4	5	5	5	4
1049		4	3	4	2	5	2
1064		4	3	5	3	5	4
1084		3	3	5	5	4	4
1104		3	4	3	3	3	2
1116		3	3	4	2	3	3
1120		3	3	4	3	4	3
SAL (*n* = 20)							
1089	Retained	4	4	1	5	0	4
1040	Delayed	3	4	4	3	2	3
1005		3	3	5	5	5	5
1014		4	4	5	5	5	4
1068		4	3	5	2	4	3
1081		3	4	5	4	5	5
1108		3	3	5	5	5	3

Note. UNGRAM = ungrammatical items; GRAM = grammatical items. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.

Summary

The results for passivization are summarized in this section. The descriptive statistics for the performances on the three measures and a visual inspection of Figures 5, 7, and 8 suggest that the four groups behaved differently on different measures. On the oral production tests, the REC+ Group showed a greater decrease of the production of indirect passives (i.e., better performance) from the posttest to the delayed test than the other groups (see Figure 5). On the elicited imitation tests, the REC and REC+ Groups showed a decrease (i.e., better judgment) from the posttest to the delayed posttest whereas the POS and SAL Groups did not (see Figure 7). On the grammaticality judgment tests, the REC+ and SAL Groups showed a greater decrease (i.e., better performance) from the posttest to the delayed posttest than the POS and REC Groups.

However, these differences were not statistically supported. Table 35 summarizes the results for the oral production tests, elicited imitation tests, and grammaticality judgment tests. Medium effect sizes were observed on the elicited imitation tests and grammaticality judgment tests. On the elicited imitation tests, the REC and SAL Groups became worse than the POS Group did. Specifically, the REC Group and the SAL Group accepted more indirect passives, whereas the POS Group did not change across the three testing times. On the grammaticality judgment tests, the POS Group became worse from the pretest-delayed posttest period whereas the SAL Group did not change.

Table 35. *Summary of Statistical Analyses for Passivization*

	Oral production test	Elicited imitation test	Grammaticality judgment test
Friedman Tests	*p* = .001	*p* = .874	*p* = .005
Post Hoc Comparison[a]			
Pretest-Posttest	*p* = .007		*p* = .005
Pretest-Delayed Posttest	*p* = .017		*p* = .002
Posttest-Delayed Posttest	*p* = .096		*p* = .936
Kruskal-Wallis Tests			
Pre-Post	*p* = .306	*p* = .047 Medium ES: REC vs. POS	*p* = .553
Pre-Delayed	*p* = .586	*p* = .113 Medium ES: POS vs. SAL	*p* = .068 Medium ES: POS vs. SAL
Post-Delayed	*p* = .215	*p* = .221	*p* = .242

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; ES = effect size; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.
[a]As a post-hoc analysis, Wilcoxon signed-rank tests were performed to compare two of the three tests.

Table 36 summarizes the results of the individual analyses for passivization. The table shows the performance across the three tests by the participants who showed (a) improvement on the posttest and maintenance on the delayed posttest (see the label Retained), (b) delayed effects, that is, improvement only on the delayed posttest (see the label Delayed), and (c) short-term effects, that is, improvement only on the posttest (see the label Short). The table also shows information about the pretest when the participants did not show improvement.

Table 36. *Summary of Individual Analyses for Passivization*

Participant	OPT	EIT	GJT
POS (*n* = 21)			
1043	C (0/4)[a]	Short	C (0/5)
REC (*n* = 22)			
1035	Delayed	B (80/75)[b]	B (4/4)[c]
REC+ (*n* = 27)			
1051	C (0/3)	Retained	Retained
1112	C (0/4)	B (100/80)	Retained
1075	C (0/4)	B (100/100)	Delayed
1049	Delayed	B (75/100)	B (4/3)
1064	Delayed	B (100/100)	B (4/3)
1119	Delayed	B (100/100)	I (3/1)
1091	Delayed	C (50/100)	C (1/4)
SAL (*n* = 20)			
1089	C (0/4)	B (75/100)	Retained
1020	Retained	B (80/100)	C (2/4)
1040	C (0/4)	B (80/100)	Delayed

Note. Retained (that is, retained effects) in the column of Effects refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the column of Effects refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the column of Effects refers to the participant who performed better on the posttest. B = acceptance of both incorrect indirect passives and correct passives. C = tendency to accept the correct passives and avoid the incorrect indirect passives, that is, correct judgment or production. I = tendency to accept the incorrect indirect passives and avoid the correct passives or tendency to accept neither of the structures, that is, unsystematic judgment or production. The shadowed cells indicate the improvement of either or both of the posttest and delayed posttest (i.e., retained effects, delayed effects, and short-term effects). POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group.
[a]The numerator indicates the production of the incorrect indirect passives whereas the denominator indicates the frequencies of the correct passives.
[b]The numerator indicates the percentage of acceptance of the incorrect indirect passives whereas the denominator indicates the percentage of acceptance of the correct passives. The percentages are used for the EIT because the numbers of items varied across the participants according to the number of correct picture selection.
[c]The numerator indicates the number of acceptance of the incorrect indirect passives whereas the denominator indicates the number of acceptance of the correct passives.

Table 36 shows that a total of 12 out of the 90 participants (13.3%; one out of 21 in the POS Group, one out of 22 in the REC Group, seven out of 27 in the REC+ Group, and three out of 20 in the SAL Group) produced or accepted both indirect passives and correct passives on at least one of the three pretests. Although the rate (13.3%) for indirect passives was smaller than that of adjective order (36.7%), the results show that several participants possessed overgeneralized rules on at least one of the three pretests.

Taking the results of the three measures together, the effects were greatest for the REC+ Group. In this group, seven out of 27 participants showed an effect. The second group was the SAL Group, three out of 20. One participant each was observed for the POS and REC Groups. However, the difference was not statistically significant: Fisher's exact test, $p = .107$.

Unlike adjective ordering, the effects seem larger on the grammaticality judgment tests and the oral production tests than the elicited imitation tests. Further, Table 36 shows that from the start, most of the participants did not use indirect passives in their production on the oral production tests; on the other hand, on the elicited imitation tests, most of the participants showed acceptance of both structures (indirect passives and correct passives) on the pretest. For indirect passives, there seemed to be no room for improvement on the oral production tests; on the elicited imitation tests, almost no improvement was observed even though there were a number of participants who overgeneralized L1 rules to English.

The findings for indirect passives are summarized as follows. The first finding concerns the first research question about the effects of exposure to no-contingent positive evidence; the other four findings are related to the other four research questions about the effects of recasts.

1. Statistical analyses show that the performance of the POS Group did not change significantly on the elicited imitation tests; on the other hand, the POS Group

changed significantly on the oral production tests and grammaticality judgment tests. More specifically, the Friedman tests and post hoc tests using Wilcoxon signed-rank tests showed that on the oral production tests, the POS Group performed significantly worse from the pretest to the posttest and from the pretest to the delayed posttest and that on the grammaticality judgment tests, the POS Group from the pretest to the posttest and from the pretest to the delayed posttest. This trend was also found for the other groups.

2. The group and individual analyses showed no statistically significant differences in the difference scores among the groups on the three measures.
3. Visual inspection of the descriptive statistics did not show clear patterns across the three tests. On the oral production tests, the REC+ Group showed a greater decrease of the production of indirect passives (i.e., better performance) from the posttest to the delayed test than the other groups. On the elicited imitation tests, the REC and REC+ Groups showed a decrease (i.e., better judgment) from the posttest to the delayed posttest; on the other hand, the POS and SAL Groups did not. On the grammaticality judgment tests, the REC+ and SAL Groups showed a greater decrease (i.e., better performance) from the posttest to the delayed posttest than the POS and REC Groups.
4. The individual analyses on the three measures indicated that more than 10% of the participants produced or accepted both the incorrect indirect passives and the correct passives.

5. In terms of effect sizes, medium effect sizes were observed on the elicited imitation tests and grammaticality judgment tests: On the elicited imitation tests, the REC Group and the SAL Group performed worse for the pretest-posttest period and the pretest-delayed posttest period, respectively, while the POS Group did not change across the testing times. On the grammaticality judgment tests, the POS Group became worse from the pretest-delayed posttest period while the SAL Group did not change.

CHAPTER 5

DISCUSSION

Research Questions/Hypotheses and Findings of This Study

The general research questions guiding the present study were: (a) Are recasts effective for narrowing overgeneralized rules in an L2, and (b) if so, how do recasts work? In order to answer these research questions, I further posited five specific research questions and corresponding hypotheses. In this section, I summarize the findings of this study in the order of the five research questions and hypotheses.

Research Question/Hypothesis 1

First, I posed the question: In relation to learning the ungrammaticality of overgeneralized rules, is exposure to non-contingent positive evidence alone effective? To this question, it was hypothesized that the POS Group would not gain from the pretest to the posttests. The findings of this study supported the hypothesis. More specifically, for both adjective ordering and indirect passives, the POS Group did not improve their performance on the three measures of implicit knowledge and explicit knowledge (oral production tests, elicited imitation tests, and grammaticality judgment tests) or even became worse from the pretest to the posttest on the oral production tests and the grammaticality judgment tests for indirect passives. Thus, the present study indicated that exposure to positive evidence did not help L2 learners learn the ungrammaticality of the color-size order and indirect passives.

Furthermore, the individual analyses on the three measures for the two structures showed that several learners produced or accepted both correct and incorrect structures. The findings suggest that the participants whose L1 was Japanese transferred L1 rules to English both implicitly and explicitly as measured by the oral production tests, elicited imitation tests, and grammaticality judgment tests.

Research Question/Hypothesis 2

The second research question asked, "in relation to learning the ungrammaticality of overgeneralized rules, is exposure to input with recasts (juxtaposed with learner utterances) more effective than exposure to non-contingent positive evidence (not juxtaposed with learner utterances)?" It was hypothesized that the REC Group would perform better than the POS Group. The findings indicated that there was a tendency that the REC Group performed better than the POS Group for adjective ordering. For example, for adjective ordering, while the POS Group did not change from the pretest to the delayed posttest, the REC Group improved from the posttest to the delayed posttest on the oral production tests and showed short-term effects from the pretest to the posttest on the other tests (see Figures 1, 3, and 4). However, these differences were small, and the statistical analyses did not yield significance in the differences between the two groups on any tests. Nevertheless, medium effect sizes for the comparison of the POS and REC Groups were obtained on the grammaticality judgment tests for the pretest-posttest period (r = .32) and for the pretest-delayed posttest period (r = .33).

On the other hand, the results for indirect passives were not clear. The REC Group showed improvement on the oral production tests and elicited imitation tests from the posttest to the delayed posttest, whereas the POS Group did not (see Figures 5 and 7); however, the difference between the two groups was not clear on the grammaticality judgment tests. A medium effect size was observed for the comparison of the POS Group and the REC Group for the pretest-posttest period on the elicited imitation tests; more specifically, the REC Group got worse while the POS Group did not change. Statistical analyses did not find significant differences between the groups.

Thus, this study found weak and partial support (i.e., for one of the two structures and one measure of the three tests) for the direct contrast hypothesis (Saxton, 1997, 2000) given that the effect sizes on the grammaticality judgment tests were medium. Below I discuss the differential effects for the two target structures and the observation of the effects on the explicit knowledge measure (i.e., grammaticality judgment tests) rather than the implicit knowledge measure (i.e., oral production tests and the elicited imitation tests).

In this section, I discuss the reasons that this study found medium effect sizes for the effects of recasts over non-contingent positive evidence on the adjective-ordering grammaticality judgment tests and that statistical analyses did not yield significance in regard to the effectiveness of recasts over non-contingent positive evidence. One possible reason is that this study was focused on syntactic rules as the target structures. Mackey, Gass, and McDonough (2000) reported that L2 learners interpreted the intention of the recasts towards lexical errors accurately (83.3% for learners of English and 66.0% for learners of Italian) whereas the interpretation of the recasts towards morpho-syntactic errors was not accurate (13.0% for learners of English and 24.0% for learners of Italian). Thus, because the target structures in this study were syntactic rules, the participants might have had difficulty in interpreting the information of recasts.

Another possible reason is that the provision of recasts with non-corrective repetition in the treatment tasks made it difficult for the participants to notice the corrective nature of the feedback. In order to control the amount of positive evidence and production opportunities, a recast or non-corrective repetition in the form of confirmation checks was provided for every utterance by the participants in the treatment tasks (see the example for the REC Group in the Procedures section). In other words, I gave confirmation checks every time the participants made an utterance. Lyster (1998a) pointed out that L2 learners might have difficulty in noticing the corrective intention of recasts on the basis of the findings that non-corrective repetition occurred in the classroom in the similar way to recasts. Likewise, recasts and non-corrective repetition provided in this study had the same communicative function (i.e., the function of confirmation checks) and had the same form of full repetition of participants' utterances. One difference between a recast and non-corrective repetition was a change of an error to a correct form.

A third possible reason is that recasts provided in this study were full recasts, that is, reformulations of all the participants' utterances. Previous studies (e.g., Ellis, Loewen, & Erlam, 2006; Sheen, 2006) reported the effectiveness of partial recasts over full recasts. For this study, full recasts were provided because the target structures in this study were syntactic rules and it was difficult to provide partial recasts, in particular, for indirect passives. Thus, the choice of full recasts might have made it difficult for the participants of this study to notice the corrective intention of recasts.

Research Question/Hypothesis 3

The third question was: In relation to learning the ungrammaticality of overgeneralized rules, is exposure to input with both recasts and additional subsequent input more effective than exposure to input with recasts? It was hypothesized that the REC+ Group would perform better than the REC Group. For adjective ordering, particularly, the REC and REC+ Groups showed similar tendencies on the elicited imitation tests and the grammaticality judgment tests (see Figures 3 and 4). Second, for indirect passives, the REC and REC+ Groups changed in a parallel manner on the elicited imitation tests; on the other hand, on the oral production tests and grammaticality judgment tests, the descriptive statistics suggest that the REC+ Group performed better than the REC Group. However, these differences were not supported by the statistical tests. In addition, no medium or large effect sizes between the two groups were found. Therefore, the results suggest that there was no difference between the REC and REC+ Groups. It must be noted that on the adjective-ordering grammaticality judgment tests, the comparison between the REC+ Group and the POS Group obtained a medium effect size (r = .32) for the pretest-delayed posttest period; the REC+ Group performed better. These findings suggest that the provision of recasts in the REC and REC+ Groups was effective at least for the adjective-ordering grammaticality judgment tests. However, no difference between the REC and REC+ Groups suggests that additional input alone might not be a necessary condition for L2 learners to retreat from the overgeneralized rules due to L1.

One possible reason for the ineffectiveness of additional input is that the frequencies of the target structures were not sufficient for the L2 learners to learn the ungrammaticality of their overgeneralized rules. In this study, the number of instances of positive evidence was controlled; both the REC and REC+ Groups received only five instances of the target structures. Long (1996) pointed out that "incorrect associations are weakened and in some cases ultimately relinquished altogether as a result of both of negative evidence and *prolonged absence of support in the input* [emphasis added]" (p. 430). Further research is needed to determine how much exposure L2 learners need to retreat from overgeneralized rules. Another possible reason is that the timing of the provision of positive evidence was inappropriate. In this study, the participants performed several tasks in a relatively short time; thus, it is possible that the participants were not provided with enough time to reflect on the interaction. According to Gass (1997), exposure to additional input after noticing an error through recasts enables L2 learners to "search for additional confirmatory or nonconfirmatory evidence" (p. 144). In other words, it is possible that L2 learners require time to process the linguistic information provided in the recasts. If this argument is correct, some interval after the treatment task in which recasts are provided might lead learners to learn the ungrammaticality of overgeneralized rules.

Research Question/Hypothesis 4

The fourth research question was: In relation to learning the ungrammaticality of overgeneralized rules, is exposure to input with enhanced salience of positive evidence (without negative evidence) more effective than exposure to input with unenhanced positive evidence (non-contingent positive evidence)? It was hypothesized that the SAL Group would perform better than the POS Group. In general, the findings of this study showed that the POS and SAL Groups changed in a similar manner on the three measures for adjective ordering (Figures 1, 3, and 4) and on the two measures, oral production tests and elicited imitation tests, for indirect passives (Figures 5 and 7) and that no statistical differences between the POS Group and the SAL Group were observed on all the measures for the two structures. On the grammaticality judgment tests for indirect passives, Figure 8 showed that the SAL Group decreased in the acceptance of ungrammatical items (i.e., improvement) from the posttest to the delayed posttest whereas the POS Group slightly increased during that period; however, the difference was not significant. Medium effect sizes were observed for the comparison between the POS Group and the SAL Group on the elicited imitation tests and the grammaticality judgment tests for the pretest-delayed posttest period. On the elicited imitation tests, the SAL Group performed worse on the delayed posttest than on the pretest, while the POS Group did not change. On the grammaticality judgment tests, the results of a Friedman test with a post hoc test using Wilcoxon signed-rank tests showed that the POS Group performed significantly worse from the pretest to the posttest and from the pretest to the posttest; on the other hand, the SAL Group did not change from the pretest (*M* = 2.60) to the delayed posttest (*M* = 2.60). These findings suggest that enhanced salience was not effective for L2 learners to learn the ungrammaticality of the overgeneralized rules about adjective ordering and indirect passives as positive evidence.

Research Question/Hypothesis 5

Lastly, I put forward the research question: In relation to learning the ungrammaticality of overgeneralized rules, does exposure to input with recasts differ from exposure to input with enhanced salience of positive evidence? It was hypothesized that no differences would be found between the REC and SAL Groups. The findings suggest that differences might have existed between the REC and SAL Groups at least for adjective ordering. For example, as mentioned above, for adjective ordering, the REC Group showed delayed effects on the oral production tests (Figure 1) and short-term effects on the elicited imitation tests and grammaticality judgment tests (Figures 3 and 4). In particular, on the grammaticality judgment tests, the comparison in the difference scores between the pretest and the delayed posttest represented a medium effect size (*r* = .37); in other words, the REC Group decreased acceptance of ungrammatical adjective order (color-size) from the pretest to the delayed posttest (the mean difference score = -0.18). On the other hand, the SAL Group did not improve on the three measures (Figures 1, 3, and 4). However, statistical analyses did not support these differences between the REC and SAL Groups. On the other hand, the results for indirect passives were mixed on the three measures; the differences between the two groups were not observed. On the oral production tests, the REC and SAL Groups changed similarly (Figure 5); on the elicited imitation tests, the REC Group showed improvement from the posttest to the delayed posttest whereas the SAL Group did not (Figure 7); and on the grammaticality judgment tests, the SAL Group improved from the posttest to the delayed test but the REC Group did not (Figure 8). The findings that the REC and SAL Groups differed for one structure and that the REC Group showed a larger effect on narrowing overgeneralized rules on the adjective-ordering grammaticality judgment tests than the POS Group suggest that enhanced salience might not play an important role in helping L2 learners narrow overgeneralized rules to more target-like ones.

Differences Between Structures

Several differences were identified between the two grammatical structures investigated in this study. First, the effects of recasts were observed for adjective ordering in terms of effect sizes. Thus, the findings of the present study lent support to the direct contrast hypothesis only for adjective ordering. Second, an inspection of the descriptive statistics suggested that the performance of the four groups on the three tests for adjective ordering was similar, whereas the patterns in the changes of the four groups differed across the three tests for indirect passives. Third, according to the individual analyses, the number and the rate of the participants who produced or accepted both correct and incorrect structures were larger for adjective ordering than those for indirect passives. In this section, possible reasons for the differences in the acquisition of the two structures are discussed.

One possibility for the differences between the two structures is the readiness of the participants in this study to acquire the two structures. As explained in the Target Structures section, the two structures differ in terms of required psycholinguistic procedures according to processability theory (Pienemann, 1998a, 1998b, 2003). In order to produce or process passivization, L2 learners need to be at an advanced level in their L2 learning. I assumed that the participants might not have difficulty processing passivization based on the results of previous studies (Izumi & Lakshmanan, 1998; Suzuki, 1996). The results of the oral production tests for indirect passives indicated that the participants in the four groups did not produce indirect passives but increased the number of indirect passives on the posttest (see Table 25 and Figure 5). In order to examine whether the participants were ready for processing passivization, I calculated the mean and standard deviations of the frequencies of correct passives on the oral production pretest. The means and standard deviations of the four groups were 3.62 and 0.59 for the POS Group, 3.36 and 1.22 for the REC Group, 2.96 and 1.29 for the REC+ Group, and 3.20 and 1.36 for the SAL Group. In other words, the participants in the four groups produced 2.96 to 3.62 instances of correct passives on the oral production pretest. Considering that the highest possible score was four, these results suggested that the participants were at the stage of L2 acquisition that would allow them to produce passivization.

A more plausible explanation of the differences between the two structures is the complexity of recasts. The recast for indirect passives contains more changes than that for adjective ordering. For this study, errors other than the target structures were reformulated in a recast as in the following example from data in this study.

Example from the Picture Identification Task (1026, REC Group)
(a) Adjective ordering

Learner:	There is green small house. [error]
Researcher:	There is a small green house? [recast]
Learner:	Yeah.
Researcher:	Number 3.

(b) Indirect passives

Learner:	Mike is broken his cup. [error]
Researcher:	Mike is sad because his cup was broken? [recast]
Learner:	Yeah.
Researcher:	Number 3.

The recast for adjective ordering (*There is a small green house?*) contained the insertion of an indefinite article as well as the change in the order of the adjectives. The recast for indirect passives (*Mike is sad because his cup was broken?*) tended to contain more changes, such as the change of a simple sentence into a complex sentence, the insertion of the conjunction *because,* the correction of the use of the tense, and the change of the subject from *Mike* to *his cup* for the predicate *was broken*. Egi (2007a) found that her participants reported interpreting recasts with more changes as responses to content. In her study, a recast with fewer changes was defined as a recast with one or two changes; a recast with more changes was referred to as a recast with three or more changes. Therefore, the differential effects of recasts for the two structures might be attributed to the differences in the number of changes in the recast.

As noted in the Theoretical Issues section, Saxton (1997) limited the direct contrast hypothesis to situations in which the learner already knows two forms that fulfill equivalent functions. For adjective ordering, only two forms fulfill equivalent functions, the color-size order and the size-color order (e.g., *a green small house* and *a small green house*). On the other hand, for indirect passives, several forms fulfill equivalent functions. An indirect passive (e.g., *Mike is broken his cup*.) can be stated as "Mike is sad because his cup was broken," "Mike's cup was broken so he became sad," "Mike is sad because someone broke his cup." It is possible that the direct contrast hypothesis does not apply to narrowing of overgeneralized rules of indirect passives because the larger number of forms that fulfill equivalent functions makes the comparison between the alternative structures more complex.

A U-shaped learning pattern was observed for indirect passives. A U-shaped behavior refers to the learning pattern that L2 learners show correct performance at first, increase incorrect performance, and finally perform correctly again (Gass & Selinker, 2001, pp. 214-215; Kellerman, 1985). On the oral production tests, the participants in the four groups produced correct passives as mentioned above and did not produce incorrect indirect passives on the pretest (see Figure 5). On the posttest, the participants in the four groups increased the production of indirect passives; on the oral production delayed posttest, they reduced the use of indirect passives. The analyses of individual performances on the oral production tests for passivization (Table 27) showed that very few participants produced both correct structures and incorrect indirect passives. Thus, it seems that the participants might have acquired the correct rules for passivization.

The results of the elicited imitation tests and grammaticality judgment tests show a similar U-shaped pattern (see Figures 7 and 8).

One possible explanation for this U-shaped pattern for indirect passives is that the provision of explanations of pictures in Japanese in the oral production tests for indirect passives might have influenced the participants' performance. The explanations in Japanese contained indirect passives (see Appendix I) and were provided on the pretest, posttest, and delayed posttest. Thus, on the later tests, the participants might have been more attentive to the explanations of pictures with the help of task familiarity and might have been induced to produce indirect passives. In other words, the participants might have used translation strategies on the oral production tests. However, because a U-shaped learning pattern was observed on the elicited imitation tests and the grammaticality judgment tests, it is difficult to argue that provision of Japanese explanations and translation strategies were the only factors causing the U-shaped pattern.

Another possible explanation is that intensive opportunities to comprehend and produce passives in the treatment tasks might have affected the participants' performance. In the treatment tasks (object placement task, picture-identification task, and picture-description task), the participants received oral input and produced passives orally. Thus, the effects of intensive use of passives might have appeared in the oral production tests more clearly than in the other tests (elicited imitation tests and grammaticality judgment tests) that required the participants to correct errors and judge the correctness of the items.

Explicit and Implicit Grammatical Knowledge

As mentioned above, in terms of effect sizes, the findings of the present study lent weak support to the direct contrast hypothesis only for grammaticality judgment tests, which were designed to tap L2 learners' explicit knowledge. In this section, I discuss the differential impacts of recasts on explicit knowledge and implicit knowledge.

On the grammaticality judgment tests, medium effect sizes were obtained for the comparison of the POS Group and the REC Group for the pretest-posttest period (r = .32) and for the pretest-delayed posttest period (r = .33). In other words, the effects of recasts appeared early on the grammaticality judgment tests. As mentioned in the Results section, an inspection of the descriptive data on the elicited imitation tests for adjective ordering suggests that the REC and REC+ Groups showed a short-term improvement (see Table 17 and Figure 3); furthermore, on the oral production tests, the REC Group showed a delayed improvement (see Table 13 and Figure 1). Based on these findings, it is possible to suppose that the impact of recasts might appear first on the measurement of explicit knowledge (i.e., grammaticality judgment tests) and then expand to the measurement of implicit knowledge (i.e., elicited imitation tests and oral production tests).

According to Long's (1996, 2007) interaction hypothesis, recasts lead L2 learners to selectively attend to target language form during communicative interactions. In other words, it is hypothesized that learners need to notice the target form for L2 acquisition. One plausible explanation for the differential effects of recasts on explicit and implicit knowledge is that noticing recasts might bring about changes in explicit knowledge more readily. Furthermore, Ellis (1991, 2004, 2005) pointed out that L2 learners rely more on explicit knowledge to judge ungrammatical sentences in a grammaticality judgment test and so argued that L2 learners use strategies for ungrammatical sentences that are different from those for grammatical sentences. Thus, the grammaticality judgment tests with ungrammatical sentences in this study detected the effects of recasts.

This interpretation seems to contradict the findings of Ellis, Loewen, and Erlam (2006) and Ellis (2007). They found improvements for the recast group and the metalinguistic feedback group on the elicited imitation tests; thus, they reported that recasts positively affected implicit knowledge. Nevertheless, they also reported that their participants performed quite well from the start on the untimed grammaticality judgment tests and metalinguistic knowledge tests. Therefore, their findings did not show that recasts lead to improvement in L2 learners' implicit knowledge without improvements in explicit knowledge; they did not contradict the above interpretation.

For indirect passivization, the results of the three tests showed a complex picture about the participants' knowledge. First, on the oral production tests, the four groups did not differ in changes of production of indirect passives (see Figure 5). As mentioned in the previous section, the participants in the four groups produced correct passives and did not produce incorrect indirect passives on the pretest. On the posttest, the participants in the four groups increased the production of indirect passives; on the oral production delayed posttest, they reduced the use of indirect passives. Second, on the elicited imitation tests, a medium effect size was obtained for the comparison of the POS and REC Groups for the pretest-posttest period: In other words, the performance of the participants in the REC Group became worse from the pretest to the posttest than the POS Group did (see Figure 7). Third, on the grammaticality judgment tests, a medium size was obtained for the comparison of the POS and SAL Groups for the pretest-delayed posttest period: In other words, the POS Group became worse from the pretest to the delayed posttest whereas the SAL Group did not change during that period (see Figure 8).

Effects of Other Factors on the Effectiveness of Recasts

Appendices S and T show the information about the learning effects on the three measures and the following factors: (a) the scores of the CELT (structure) test (the CELT column), (b) the number of recasts provided in the treatment tasks for the REC and REC+ Groups (the REC column), (c) the knowledge of the target structure before the present study (the PREV column), (d) the scores of the meta-language test (the MLT column), and (e) the degree of awareness that the purpose of the study was to examine the knowledge of the target structure (the EXIT column).

First, the effects of the CELT scores on the effectiveness of recasts are discussed. According to Appendices S and T, there seems to be no clear impact of the CELT scores on the learning effectiveness. Among the participants who showed learning effects (see the Retained, Delayed, and Short in the OPT, EIT, GJT columns), the CELT scores ranged from 10 (1093 in the REC Group) to 67 (1114 in the REC+ Group) for adjective ordering (see Appendix S) and from 23 (1119 in the REC+ Group) to 59 (1075 in the REC+ Group) for indirect passives. I also examined correlations between the CELT scores and the difference scores between two of the pretest, posttest, and the delayed posttests for each target structure. Because the distributions of the difference scores deviated from the normality as mentioned in the Results chapter, I calculated Kendall's tau-b correlation coefficients (see Appendices U and V). For adjective ordering, of the 36 correlations, the probabilities of three coefficients were less than .05: for the pretest-posttest period on the elicited imitation tests (the REC Group), the pretest-delayed posttest period on the elicited imitation tests (the REC+ Group), and the pretest-posttest period on the grammaticality judgment tests (the REC Group). For indirect passives, of the 36 correlations, the probabilities of two coefficients were less than .05: for the pretest-posttest period on the elicited imitation tests. These two correlations were derived from the POS and SAL Groups. Taken together, the participants in the REC Group who got higher CELT scores tended to improve from the pretest to the posttest on the elicited imitation tests and grammaticality judgment tests; on the other hand, the participants in the REC+ Group who got higher CELT scores tended to improve from the pretest to the delayed posttest on the elicited imitation tests. However, these changes were not reflected in learning effects as shown in the individual analyses. In contrast, for indirect passives, the CELT scores seemed to have no impact on the improvements from the pretests to the posttests or on the learning effects.

Second, the effects of the numbers of recasts are discussed. In this study, the average numbers of recasts provided to the REC and REC+ Groups were 1.85 and 2.06

for adjective ordering and 3.05 and 3.04 for indirect passives, respectively (see Table 12 in the Preliminary Analyses section). The treatment task during which recasts were provided took about five minutes. The numbers of recasts were small in comparison with previous studies (Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Loewen & Nabei, 2007). For example, the recast group in Ellis (2007) received 42 recasts for the past tense and 22 recasts for the comparative during the one-hour tasks lasting for two days. In Loewen and Nabei's study, 18.7 recasts were provided during the 30-minute tasks. Thus, it is plausible that if a greater number of recasts had been provided, the effects would have been large enough to reach statistical significance.

However, a closer look at the results of this study suggests that the number of recasts might not have influenced subsequent learning. First, the present study showed that the effects of recasts were observed on the grammaticality judgment tests for adjective ordering, but no effects were found on any measures for indirect passives even though more recasts were provided for indirect passives than for adjective ordering. Second, even one recast sometimes led to improvement (see Appendices S and T). For example, for adjective ordering, the number of recasts provided to the participants who showed improvements on the two posttests (see the label Retained in Appendix S) ranged from one (for 1034 and 1007) to five (for 1083) for the REC Group and from one (for 1112 and 1114) to three (for 1017) for the REC+ Group. For indirect passives, two participants in the REC+ Group improved on the two posttests (see the label Retained in Appendix T); these two participants received two recasts, which was less than the average number (3.04). In addition, I examined the correlations between the numbers of recasts and the difference scores between two of the pretest, posttest, and delayed posttest (see Appendix W). Kendall's tau-b correlation coefficients were calculated because the distributions of the difference scores deviated from normality as mentioned in the Results chapter. For adjective ordering, of the 18 correlations, probabilities of two coefficients derived from the REC+ Group were less than .05; for indirect passives, of the 18 correlations, the probabilities of two coefficients were less than .05 for the pretest-posttest period on the oral production tests (the REC and the REC+ Group). However, three of these four coefficients were positive: In other words, a larger number of recasts led to more acceptance or production of incorrect structures. Only for the pretest-posttest period, participants in the REC+ Group who received more recasts performed better on the grammaticality judgment tests. Taken together, these findings suggest that the number of recasts is not an influential factor in acquisition.

Regarding the frequency of recasts, it must be pointed out that although the way that recasts were provided in this study was unnatural (i.e., the recasts were provided every time errors were made in a laboratory setting), the frequencies of the recasts during the treatment tasks were similar to naturalistic environments. In an observation study of naturalistic interaction in the classroom, Lyster and Ranta (1997) reported the numbers of recasts produced by four teachers: 93 for T3, 96 for T4, 116 for T5, and 70 for T6 (p. 53). They also reported the observation time for each lesson: 340 minutes for T3, 250 minutes for T4, 325 minutes for T5, and 185 minutes for T6 (p. 62). Based on these results, the numbers of recasts per five minutes as in the present study were calculated as follows: 1.37 for T3, 1.92 for T4, 1.78 for T5, and 1.89 for T6. Therefore, the number of recasts provided in this study, in particular those for adjective ordering, were not so different from the provision of recasts by the teachers studied by Lyster and Ranta. Mackey (2012) pointed out that the intensive provision of recasts in laboratories loses external validity because they do not take place in naturalistic environments; however, this experimental study has some external validity in terms of the frequency of recasts.

Third, the effects of the knowledge of the target structure before the present study on the effectiveness of recasts are discussed. In the follow-up questionnaire (Appendix C), I provided the rules that the color-size-noun sequence and indirect passives are ungrammatical in English and asked whether participants had known the rules. I assigned 2 points to responses indicating they had known the rules, 1 point to responses indicating that they might have known the rules, and 0 points to responses indicating that they had not known the rules. I then calculated Kendall's tau-b correlation coefficients (see Appendices X and Y). For adjective ordering, of the 36 correlations, only one coefficient showed a probability of less than .05 for the posttest-delayed posttest period on the grammaticality judgment tests. This coefficient was derived from the REC+ Group. For indirect passives, of the 36 correlations, probabilities of five coefficients were less than .05: for the pretest-posttest period on the oral production tests (the POS and REC+ Groups), for the pretest-delayed posttest period on the oral production tests (the POS Group), for the pretest-posttest period on the elicited imitation tests (the REC+ Group), and for the pretest-delayed posttest period on the elicited imitation tests (the REC+ Group). Thus, previous knowledge of the target structures seemed to have impact on the improvements for the REC+ Group, in particular, for indirect passives. However, such effects of previous knowledge were not observed for the REC Group.

I have discussed the factors of the CELT scores, the number of recasts, and the previous knowledge on the target structures. In general, these factors did not exert a clear impact on the improvements or the learning effects. However, the findings that several correlations between these factors and the difference scores were high suggest that in the future, it is necessary to examine the effects of these factors on the effectiveness of recasts in more controlled designs.

Kruskal-Wallis tests were performed to examine whether the four groups differed in the scores of the meta-language test (the MLT column) and the degree of awareness that the purpose of the study was to examine the knowledge of the target structure (the EXIT column). For adjective ordering, the tests showed no significant differences between the groups: $\chi^2(3) = 4.24$, $p = .237$ for the pretest-posttest period; $\chi^2(3) = 0.41$, $p = .938$ for the meta-language test and $\chi^2(3) = 3.82$, $p = .281$ for the awareness of the purpose of the study. For indirect passives, the tests showed no significant differences between the groups either: $\chi^2(3) = 4.01$, $p = .260$ for the meta-language test and $\chi^2(3) = 5.17$, $p = .160$ for the awareness of the purpose of the study. In other words, the differences in the conditions among the four groups did not have an impact on the scores of the meta-language test and the awareness that the purpose of the study was to examine adjective ordering or passivization.

Theoretical Implications

The findings of this study have several theoretical implications. The first implication concerns the influence of L1 in the L2 learners' interlanguage and the role of positive evidence in learning of the ungrammaticality of overgeneralized rules. In regard to the first research question, the findings that several participants produced or accepted both correct and incorrect structures on the three measures for the two structures suggest that because of L1 influence, L2 learners possess overgeneralized rules that allow correct and incorrect structures. Furthermore, the findings that the POS Group did not improve from the pretest to the posttests on either of the three measures for adjective ordering or for indirect passives suggest that L2 learners have difficulty in narrowing the overgeneralized rules on the basis of positive evidence. In the present study, the POS Group was provided with 10 instances of the correct target structures in the communicative tasks (the object placement task and the picture identification task). Thus, the findings suggest that ten tokens of the correct size-color-noun sequence and correct passives do not work as indirect negative evidence: In other words, L2 learners of English in this study did not learn the ungrammaticality of the color-size-noun sequence or indirect passives on the basis of the observation of the non-occurrence of the incorrect structures in the input. Taken together, the results support White's (1987, 1989a) argument that positive evidence might not be sufficient for L2 learners to retreat from overgeneralization about L2 rules due to L1 transfer. The findings also support previous empirical studies (Izumi & Lakshmanan, 1998; Sakai, 2003; Trahey, 1996;

Trahey & White, 1993; White, 1989b, 1991) showing that positive evidence alone does not lead to learning of the ungrammaticality of overgeneralized forms.

The second theoretical implication concerns the following two theoretical hypotheses concerning the mechanisms of recasts as negative evidence: the direct contrast hypothesis and the additional input hypothesis. The findings related to the second and third research questions showed that the REC Group improved better than the POS Group on the adjective-ordering grammaticality judgment tests and that the comparison between the REC and POS Groups yielded medium effect sizes although statistical significance was not obtained. The findings also showed that in terms of statistical analysis and effect sizes, the REC and REC+ Groups did not differ on the three measures for the two structures. These findings suggest that the direct contrast hypothesis (Saxton, 1997, 2000) was weakly and partially supported, for one structure, adjective ordering, in terms of effect sizes; on the other hand, the additional input hypothesis (Gass, 1997; Gass & Mackey, 2007; Long, 1996, 2007) was not clearly supported. According to the direct contrast hypothesis, the contingency and juxtaposition of the correct reformulation with the learner's erroneous utterance enable the learner to notice the difference between the two. On the other hand, according to the additional input hypothesis, L2 learners need not only to notice the error but also to search for further evidence in the input for internalization of the target form. The findings of this study suggest that the contingency and juxtaposition of the erroneous utterance and the correct reformulation are necessary conditions for L2 learning and that additional input provided after the provision of recasts might not be necessary, even though additional input might facilitate learning.

The third theoretical implication concerns the following theoretical hypothesis concerning the mechanisms of recasts as negative evidence: the enhanced salience hypothesis. The findings related to the fourth research question showed that the SAL and POS Groups did not differ in terms of statistical analyses and effect sizes except for the pretest-delayed posttest period on the elicited imitation tests and grammaticality judgment tests. The effect sizes for the pretest-delayed posttest period on the elicited imitation tests and grammaticality judgment tests were medium, but the results showed that the SAL Group did not improve on the tests. The findings related to the fifth research question showed that the REC and SAL Groups differed for adjective ordering. These findings suggest that Leeman's (2003) claim that recasts are effective not because they provide negative evidence but because they provide input with enhanced salience due to negative evidence must be considered to be limited to the learning of non-salient forms. In other words, Leeman's enhanced salience hypothesis can be interpreted to be limited to learning some structures or forms, not to concern retreat from overgeneralized rules.

In sum, the findings of the study imply that contingency and juxtaposition might be necessary conditions for recasts to function as negative evidence and that additional input might be a facilitating condition. The roles for salience are not clear. Table 37 presents the implications of the findings of the study.

Table 37. *Examination of the Three Hypotheses*

Hypothesis	Condition	Findings
Direct Contrast Hypothesis	Contingency and juxtaposition	Partially supported
Additional Input Hypothesis	Additional input	Not supported
Enhanced Salience Hypothesis	Enhanced salience	Not supported

Pedagogical Implications

This study provides several pedagogical implications. First, the findings of this study that ten tokens of positive evidence did not lead to statistically significant improvements in the POS Groups suggest that Japanese learners of English need more than positive evidence to learn the ungrammaticality of the color-size-noun sequence and indirect passives in English. In other words, this study identified the color-size-noun sequence and indirect passives as grammatical areas that Japanese learners of English need negative evidence. Several researchers (e.g., Gass, 1997; Long, 1996; White, 1987, 1989a) have pointed out that when the learner's interlanguage rules allow both incorrect and correct forms, the learner needs negative evidence to narrow the overgeneralized rules.

Second, teachers can utilize recasts as one option of negative evidence. Although statistical significance was not obtained, the medium effect sizes suggest that recasts helped Japanese learners of English learn the ungrammaticality of the color-size-noun sequence in English. In other words, the findings of this study show that recasts provide L2 learners with negative evidence and facilitate L2 learning. According to Long and Robinson (1998), negative evidence can be provided preemptively as in the form of grammar rules or reactively as in the form of negative feedback. Recasts are one implicit form of negative feedback. Because the implicitness of recasts might not interrupt the flow of communication, recasts should be utilized as an option of the forms of negative evidence in communication-oriented L2 classrooms. However, unlike full recasts provided in this study, partial/reduced recasts are recommended.

Third, as mentioned above, the findings of this study showed that the exposure to non-contingent positive evidence did not work for L2 learners to learn the ungrammaticality of the overgeneralized rules about adjective ordering and passivization. The findings provide suggestions about the appropriate moment to provide feedback. In other words, teachers should not miss the moment at which L2 learners make errors and should provide recasts or other forms of negative feedback.

Fourth, in this study, additional input after the provision of recasts was found to be ineffective for learning the ungrammaticality of the color-size-noun sequence and indirect passives. The findings suggest that teachers do not have to ensure additional input after the provision of recasts. It is observed very often that the discourse in the L2 classroom becomes unnatural because teachers use certain structures repeatedly: For example, teachers only produce the progressive *–ing* in their talk in one lesson in order to provide sufficient amount of positive evidence of the progressive *–ing*. However, this study suggests that such input enhancement might be helpful but unnecessary for L2 learners to learn the ungrammaticality of overgeneralized rules. Rather, as pointed out above, it might be more effective to provide recasts at the right moment.

CHAPTER 6

CONCLUSION

Summary of the Findings

The findings of this study can be summarized as follows:

1. The POS Group performed neither better nor worse on the adjective ordering posttests or the delayed posttests of the three measurements. For indirect passives, the performance of the POS Group did not change significantly on the elicited imitation tests; on the other hand, the POS Group performed statistically significantly worse on the oral production and grammaticality judgment posttests and delayed posttests.
2. For adjective ordering, an inspection of the descriptive statistics showed the following patterns. The REC Group performed better on the oral production delayed posttest than on the posttest, whereas the other groups did not change or performed worse. On the elicited imitation tests, the REC and REC+ Groups showed a short-term improvement although the effects disappeared on the delayed posttest. The REC and REC+ Groups improved on the grammaticality judgment posttest and maintained the improvements on the delayed test. For indirect passives, an inspection of the descriptive statistics did not show clear patterns across the three tests. On the oral production tests, the REC+ Group showed a greater decrease in the production of indirect passives (i.e., better performance) from the posttest to the delayed test than the other groups. On the elicited imitation tests, the REC and REC+ Groups

showed a decrease (i.e., better judgment) from the posttest to the delayed posttest; on the other hand, the POS and SAL Groups did not. On the grammaticality judgment tests, the REC+ and SAL Groups showed a greater decrease (i.e., better performance) from the posttest to the delayed posttest than the POS and REC Groups.

3. For the two target structures, both the group analysis and individual analysis showed no statistically significant differences among the four groups on the three measures (oral production tests, elicited imitation tests, and grammaticality judgment tests).
4. Medium effect sizes were observed on the grammaticality judgment tests for adjective ordering and on the elicited imitation tests for indirect passives. More specifically, for adjective ordering, the improvement of the REC Group from the pretest to the posttest was larger than that of the POS Group, and the improvements of the REC and REC+ Groups were larger than those of the POS and SAL Groups from the pretest to the delayed posttests. For indirect passives, the REC Group and the SAL Group performed worse for the pretest-posttest period and the pretest-delayed posttest period respectively while the POS Group did not change across the testing times.

The findings suggest that L2 learners might persist in using incorrect overgeneralized rules due to L1 morpho-syntax and that positive evidence is not effective in helping L2 learners retreat from the overgeneralized rules.

Two research questions guided the present study: "Are recasts effective for narrowing overgeneralized rules in an L2?" and "If so, how do recasts work?" The answer obtained through this study is that recasts are effective for narrowing *some* overgeneralized L2 rules. The nature of contingency and juxtaposition might be necessary conditions for recasts to provide L2 learners with negative evidence and facilitate L2 learners' retreat from overgeneralized rules due to L1. In other words, this study lent weak and partial support (i.e., for adjective ordering) to the direct contrast hypothesis (Saxton, 1997, 2000) as a mechanism underlying recasts.

Limitations

One limitation of this study was the sample size. This study did not produce statistically significant results in the group analyses and individual analyses although some comparisons of the groups yielded medium effect sizes. One plausible reason is that the sample size for this study was too small to detect statistical significance. Using the software G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2014), the necessary sample size was calculated under the following conditions: power = 0.80, effect size = 0.52 (transformed from r, .32 for the comparison of the POS Group and the REC Group for the pretest-delayed posttest period on the grammaticality judgment test, to Cohen's d), and probability = .050. The required sample size for each group is 49. The sample size of both the POS Group and the REC Group in this study was 20, smaller than the required sample size.

Another limitation, closely related to the first limitation, was that the power analysis was not performed before designing and conducting the study. Doing the power analysis might have increased the possibility of obtaining significance in statistical analyses.

A third limitation was the small number of recasts provided in the treatment tasks. As Long, Inagaki, and Ortega (1998, p. 368) pointed out, the small number of recasts has the risk of underestimating the effectiveness of recasts. In the previous experimental studies that reported the effectiveness of recasts, the number of recasts provided to individual participants ranged from six (Long, Inagaki, & Ortega, 1998) to 24 (Carroll & Swain, 1993). In Ellis (2007), 42 recasts (to the past tense errors) and 22 recasts (to the comparative errors) were provided to the experimental group in a classroom setting; in Loewen and Nabei (2007) whose target structure was question-formation in English, 18.7 recasts on average were provided to a group of four participants. As discussed in the Differences Between Structures section, the number of changes in the recast for indirect passives tends to be large. Previous studies have shown that the large number of changes might affect learner's noticing of errors or uptake (Doughty, 1994; Egi, 2007a; Philp, 2003). Thus, it is possible to argue that L2 learners need to receive more recasts in order to process the information provided in recasts, in particular, for complex structures like indirect passives. As pointed out in the section of Effects of Other Factors on the Effectiveness of Recasts, several participants benefitted even from one recast; nevertheless, if more recasts had been provided in this study, the effectiveness of recasts would have been more explicit, and statistical analyses might have yielded significance.

Fourth, in this study, the CELT scores were not used for the group assignment although I examined the equivalence of the four groups in terms of grammatical knowledge on the basis of the CELT scores. As mentioned in the Discussion section, the results of this study suggested that there were differential effects of the CELT scores on the improvements on the tests between adjective ordering and indirect passives. If the participants had been screened on the basis of the CELT scores, statistical analyses might have obtained significance particularly for adjective ordering.

Suggestions for Future Research

Several recommendations for future research can be made on the basis of the present study. First, the type of recasts should be considered for future studies. This study was focused on one type of recasts, that is, full repetition of an L2 learner's erroneous utterance with correction. This approach might lessen the salience of a recast for learners. Egi (2007a) showed that the length of a recast influenced L2 learners' interpretation of a recast as negative feedback. As reviewed in the Definitional Issues section, recasts have been defined differently in studies. For example, Ellis, Loewen, and Erlam (2006) and Ellis (2007) included partial recasts in the recast condition and reported that recasts and metalinguistic feedback were more effective than no-provision of feedback. On the other hand, Loewen and Nabei (2007) utilized full-length recasts. It is plausible that with partial recasts, the effects of recasts would have been stronger. Further studies with different lengths of recasts are needed to determine the effects of recasts as negative evidence that lead to a narrowing of overgeneralized rules to more target-like ones.

Another suggestion is to ensure triangulation for data gathering. Following Ellis's (2005) suggestion, in this study, three types of testing instruments were utilized in order to tap both explicit and implicit L2 knowledge. The findings suggest that effects of the treatment appeared differently with different measures. Thus, it is possible that different results will be obtained with different measures. The findings of this study supported Ellis's (2005) suggestion that instruments designed to measure both explicit and implicit grammatical knowledge should be needed. Furthermore, several testing instruments to examine either explicit or implicit grammatical knowledge will be needed in future research to increase the reliability of the findings. In this study, oral production tests and elicited imitation tests were used to measure implicit grammatical knowledge whereas only a grammaticality judgment test was utilized to measure explicit knowledge, and medium effect sizes for the effects of treatments were obtained only on the adjective-ordering grammaticality judgment tests. The findings would be more reliable with triangulation.

Lastly, future researchers should examine learners' perceptions and interpretations of recasts in terms of the ungrammaticality of overgeneralized rules. Egi (2007a) and Sakai (2004, 2011) used stimulated recall techniques to investigate learners' perceptions and interpretations of recasts and showed that L2 learners received negative evidence from recasts; however, they did not investigate errors that L2 learners might possibly reduce with the exposure to positive evidence. Researchers need to examine whether L2 learners interpret recasts as negative evidence and utilize such information for narrowing overgeneralized rules.

Final Conclusions

The findings of the present study are of theoretical and practical significance. Theoretically, this study showed empirical evidence that recasts provided negative evidence to L2 learners and that L2 learners can narrow overgeneralized rules due to L1 transfer to more target-like rules, although the effects of recasts were limited to the grammaticality judgment tests for adjective ordering. Previous studies (e.g., Egi, 2007a; Sakai, 2004, 2011) showed that L2 learners interpret recasts as negative evidence and utilize the linguistic information in their immediate responses. This study moves the body of research knowledge forward by showing that recasts lead to changes in L2 knowledge, minimally where explicit knowledge is concerned.

In addition, three hypotheses—the direct contrast hypothesis, the additional input hypothesis, and the enhanced salience hypothesis—and the mechanisms underlying recasts were examined in this study. The findings lent support to the direct contrast hypothesis only. Although further research is needed if generalizations are to be made, I believe that this study is a first step toward understanding the mechanisms underlying recasts.

Practically, the findings of this study showed that incorrect adjective ordering and indirect passives are not eliminated only with positive evidence; thus, negative evidence is necessary to narrow overgeneralized rules due to L1 syntactic structures. Second, it was found that recasts effectively helped L2 learners retreat from overgeneralized rules, in particular, for adjective ordering. Recasts can be a useful type of feedback in communicative language classrooms because they do not interfere with the flow of communication.

A great number of researchers have already examined the effects of recasts; however, in order to fully understand the mechanisms underlying recasts, further research is needed.

REFERENCES

Abe, Y. (1999). Bun no shikumi [Sentence structures]. In K. Inoue, K. Harada, & Y. Abe (Eds.), *Seisei gengogaku nyuumon* [Introduction to generative linguistics] (pp. 207-251). Tokyo, Japan: Taishukan.

Arthur, B. (1980). Gauging the boundaries of second language competence: A study of learner judgments. *Language Learning, 30,* 177-194.

Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of the passé composé and imparfait. *The Modern Language Journal, 85,* 226-243.

Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study. *The Modern Language Journal, 88,* 31-55.

Berent, G. P. (1994). The subset principle in second-language acquisition. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 17-39). Hillsdale, NJ: Erlbaum.

Bialystok, E. (1979). Explicit and implicit judgments of L2 grammaticality. *Language Learning, 29,* 81-103.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245-261). Hillsdale, NJ: Erlbaum.

Bohannon, J. N., Padgett, R. J., Nelson, K. E., & Mark, M. (1996). Useful evidence on negative evidence. *Developmental Psychology, 32,* 551–555.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15,* 357-386.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston, MA: Heinle & Heinle.

Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning, 27*, 29-46.

Chaudron, C. (1983). Research on metalinguistic judgments: A review of theory, methods, and results. *Language Learning, 33,* 343-377.

Close, R. A. (1975). *A reference grammar for students of English.* Harlow, England: Longman.

Cobb, T. (2006). Web Vocabprofiler (Version 2.7) [Computer software]. Retrieved from http://www.lextutor.ca/vp/eng/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Erlbaum.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 2,* 213-238.

Declerck, R. (1991). *A comprehensive descriptive grammar of English.* Tokyo, Japan: Kaitakusha.

Doughty, C. (1994). Fine-tuning of feedback by competent speakers to language learners. In J. Alatis (Ed.), *GURT 1993: Strategic interaction and language acquisition* (pp. 96-108). Washington D.C.: Georgetown University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York, NY: Cambridge University Press.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two.* Oxford, England: Oxford University Press.

Eckman, F. R. (1994). Local and long-distance anaphora in second-language acquisition. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 207-225). Hillsdale, NJ: Erlbaum.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness, 13,* 243-264.

Egi, T. (2007a). Interpreting recasts as linguistic evidence: The roles of linguistic target, length, and degree of change. *Studies in Second Language Acquisition, 29*, 511-537. doi: 10+10170S0272263107070416

Egi, T. (2007b). Recasts, learners' interpretations, and L2 development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 249-267). Oxford, England: Oxford University Press.

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *The Modern Language Journal, 94,* 1-21. doi: 10.1111/j.1540-4781.2009.00980.x

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition, 13,* 161-186.

Ellis, R. (1994). A theory of instructed second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79-114). San Diego, CA: Academic Press.

Ellis, R. (2004). The definition and measurement of explicit knowledge. *Language Learning, 54,* 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27,* 141-172. doi: 10+10170S0272263105050096

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 339-360). Oxford, England: Oxford University Press.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28,* 339-368. doi: 10+10170S0272263106060141

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27,* 464-491. doi:10.1093/applin/aml001

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2014). G*Power 3.1.9.2 [Computer software]. Retrieved from http://www.gpower.hhu.de/

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Gass, S. (1983). The development of L2 intuitions. *TESOL Quarterly, 17,* 273-291.

Gass, S. M. (1994). The reliability of second-language grammaticality judgments. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 303-322). Hillsdale, NJ: Erlbaum.

Gass, S. M. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Erlbaum.

Gass, S. M., & Lewis, K. (2007). Perceptions about interactional feedback: Differences between heritage language learners and non-heritage language learners. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 79-99). Oxford, England: Oxford University Press.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Erlbaum.

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 175-199). Mahwah, NJ: Erlbaum.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Mahwah, NJ: Erlbaum.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly, 36,* 543-572.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2,* 1-23.

Hara, M., & Ma, C-H. (2004). Why do L2 learners optionally choose a certain divergent analysis of TL over a TL-like one? *EUROSLA Yearbook, 4,* 7-40.

Harris, D. P., & Palmer, L. A. (1986). *A comprehensive English language test for learners of English: Form A test book*. New York, NY: McGraw-Hill.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form *–te i- (ru)* by learners of Japanese as a foreign language. *Language Learning, 54,* 311-394.

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition, 25,* 1-36. doi:10.1017.S0272263103000019

Izumi, S., & Lakshmanan, U. (1998). Learnability, negative evidence and the L2 acquisition of the English passive. *Second Language Research, 14,* 62-101.

Johnson, Y. (2008). *Fundamentals of Japanese grammar: Comprehensive acquisition.* Honolulu, HI: University of Hawai‘i Press.

Kellerman, E. (1985). If at first you *do* succeed... In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 345-353). Rowley, MA: Newbury House.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Oxford, England: Pergamon.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom.* Oxford, England: Pergamon.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition, 25,* 37-63. doi:10.1017. S0272263103000020

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. In R. DeKeyser (Ed.), *Practice in second language learning: Perspectives from linguistics and psychology.* Cambridge, England: Cambridge University Press.

Littlewood, W. (1981). *Communicative language teaching.* Cambridge, England: Cambridge University Press.

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 361-377). Oxford, England: Oxford University Press.

Long, M. H. (1980). *Input, interaction, and second language acquisition.* (Unpublished doctoral dissertation). University of California, Los Angeles.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. B. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-51). Amsterdam, The Netherlands: Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Long, M. H. (2007). *Problems in SLA.* Mahwah, NJ: Erlbaum.

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal, 82,* 357-371.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge, England: Cambridge University Press.

Lyster, R. (1998a). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20,* 51-81.

Lyster, R. (1998b). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48,* 183-218.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37-66.

Mackey, A. (1995). *Stepping up the pace: Input, interaction and interlanguage development: An empirical study of questions in ESL.* (Unpublished doctoral dissertation). University of Sydney, Australia.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21,* 557-587.

Mackey, A. (2007). Introduction: The role of conversational interaction in second language acquisition. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 1-26). Oxford, England: Oxford University Press.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning.* Oxford, England: Oxford University Press.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22,* 471-497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 407-462). Oxford, England: Oxford University Press.

Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System, 30,* 459-477.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82,* 338-356.

Ministry of Education, Culture, Sports, Science and Technology. (2001). *Practical handbook for elementary school English activities.* Tokyo, Japan: Kairyudo.

Ministry of Education, Culture, Sports, Science and Technology. (2008). *Chuugakkou gakushu shidouyouryou kaisetsu gaikokugo hen* [Junior high school teaching guide for the course of study: Foreign languages]. Tokyo, Japan: Kairyudo.

Ministry of Education, Culture, Sports, Science and Technology. (2010). *Koutougakkou gakushu shidouyouryou kaisetsu gaikokugo hen eigo hen* [Senior high school teaching guide for the course of study: Foreign languages and English]. Tokyo, Japan: Kairyudo.

Morris, F. A., & Tarone, E. E. (2003). Impact of classroom dynamics on the effectiveness of recasts in second language acquisition. *Language Learning, 53,* 325-368.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 227-243). Hillsdale, NJ: Erlbaum.

Nabei, T., & Swain, M. (2002). Learner awareness of recasts in classroom interaction: A case study of an adult EFL student's second language learning. *Language Awareness, 11,* 43-63.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51,* 719-758.

Oliver, R. (1995). Negative feedback in child NS/NNS conversation. *Studies in Second Language Acquisition, 17,* 459-481.

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning, 50,* 119-151.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-420). San Diego, CA: Academic Press.

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25,* 99-126. doi: 10.1017.S0272263103000044

Pienemann, M. (1998a). *Language processing and second language development: Processability theory.* Amsterdam, Holland: Benjamins.

Pienemann, M. (1998b). Developmental dynamics in L1 and L2 acquisition: Processability theory and generative entrenchment. *Bilingualism: Language and Cognition, 1,* 1-20.

Pienemann, M. (2003). Language processing capacity. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 679-714). Oxford, England: Blackwell.

Pienemann, M., Di Biase, B., & Kawaguchi, S. (2005). Extending processability theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of processability theory* (pp. 199-251). Amsterdam, Holland: Benjamins.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge, England: Cambridge University Press.

Rivers, W. M. (1968). *Teaching foreign-language skills.* Chicago, IL: The University of Chicago Press.

Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. Schmidt (Ed.), *Attention & awareness in foreign language learning* (pp. 163-182). Honolulu, HI: University of Hawaii at Manoa.

Sakai, H. (2003). Gojunno gakushuuniokeru kansetsuhiteishoukono yakuwari: kouteishoukowa hibunpouseinogakushuuni koukaga aruka [Can positive evidence lead to recognition of ungrammaticality?]. *Komyunikeishonto Gengokyouiku (Shinshu University Research on Communication and Language Education, SURCLE), 3*, 27-37. Retrieved from http://snow.shinshu-u.ac.jp/~surcle/surcle03/sakai.pdf

Sakai, H. (2004). Roles of output and feedback for L2 learners' noticing. *JALT Journal, 26,* 25-54.

Sakai, H. (2008). An analysis of Japanese university students' oral performance in English using processability theory. *System, 36,* 534-549. doi:10.1016/j.system.2008.03.002

Sakai, H. (2011). Do recasts promote noticing the gap in L2 learning? *The Asian EFL Journal Quarterly, 13,* 377-410.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24*, 139-161.

Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language, 20,* 221-252. doi:10.1177 014272370002006001

Schachter, J. (1991). Corrective feedback in historical perspective. *Second Language Research, 7*, 89-102.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting *competence* and *linguistic behavior*. *Studies in Second Language Acquisition, 15,* 147-163.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*, 361-392. doi:10.1191/1362168806lr203oa

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375-409). San Diego, CA: Academic Press.

Spada, N., & Fröhlich, M. (1995). *COLT observation scheme: Coding conventions and applications.* Sydney: National Centre for English Language Teaching and Research (NCELTR), Macquarie University.

Suzuki, H. (1996). *The transfer of perspective in the indirect passive.* Paper presented at the Annual Meeting of the International Conference on Pragmatics and Language Learning. (ERIC Document Reproduction Service No. ED 402793)

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 51,* 158-164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford, England: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge, England: Cambridge University Press.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal, 82,* 320-337.

Swan, M. (1995). *Practical English usage* (2nd ed.). Oxford, England: Oxford University Press.

Thomson, A. J., & Martinet, A. V. (1986). *A practical English grammar* (4th ed.). Oxford, England: Oxford University Press.

Trahey, M. (1996). Positive evidence in second language acquisition: Some long-term effects. *Second Language Research, 12,* 111-139.

Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition, 15,* 181-204.

Van den Branden, K. (1997). Effects of negotiation on language learners' output. *Language Learning, 47,* 589-636.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence. *Applied Linguistics, 8*, 95-110.

White, L. (1989a). *Universal grammar and second language acquisition.* Amsterdam, Holland: Benjamins.

White, L. (1989b). The adjacency condition on case assignment: Do L2 learners observe the subset principle? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 134-158). Cambridge, England: Cambridge University Press.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7,* 133-161.

White, L. (2007). Linguistic theory, universal grammar, and second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 37-55). Mahwah, NJ: Erlbaum.

Yasui, M. (1996). *Kaiteiban eibunpo soran* [A better guide to English grammar]. Tokyo, Japan: Kaitakusha.

APPENDICES

APPENDIX A

CONSENT FORM

研究参加の同意書

研究の目的

　ご協力ありがとうございます。本研究の目的は、英語の熟達度（スピーキング、文法力など）と英語によるインタラクションの関係を調べることです。結果は、第二言語としての英語習得のメカニズムの解明や、第二言語としての英語教授法の示唆に貢献すると考えています。

方法

　3 度のセッションがあります。1 回目は 90 分、2 回目と 3 回目は 60 分間程度です。意識過剰にならず、自分の力で、英語で言えることを表現してください。

研究協力について

　この研究に協力するかどうかは任意です。強制いたしません。研究協力に同意しないことで、対象者に不利な事態が生ずることはありません。

　セッションを途中で止める権利があります。また、収集されたデータを研究に使用するか判断することができます。セッションを途中で止めた場合や、研究に使用してもらいたくないという希望の場合は、データを廃棄します。

　3 度のセッションを完了した方には、謝礼として 2,000 円お支払いいたします。

個人情報の保護

　分析の段階及び研究成果の発表の段階で、個人が特定できないように配慮します。そのために、希望する偽名（名字と名前）を記入してください（研究発表の際には、偽名を使います）。ただし、男性は典型的な男性の名前を、女性は典型的な女性の名前をつけてください。例えば、次のようになります。（本名　酒井　英樹　偽名　信州太郎　論文中 Taro）

研究成果の発表について

　収集されたデータ（録音された音声データ、録画された画像データ、テストの回答、アンケートの回答）は、学会発表や研究論文として発表されることがあります。その場合、個人が特定できないように配慮します。

　以上の研究の主旨を理解し、研究に参加してくださる場合には、署名してください。

日付	署名
______________________________（年月日）	____________________（姓名、本名）

APPENDIX B

BACKGROUND QUESTIONNAIRE

アンケート（事前用）

あなたについて教えてください。

(1) 日付 2012 年（ ）月（ ）日

(2) 氏名（ ） 偽名（苗字と名前）（ ）
※男性は男性らしい偽名、女性は女性らしい偽名

(3) 性別 （○をしてください） 男性 女性

(4) 年齢 （ 歳 ヶ月）（実験日現在）

(5) 国籍（○をしてください） 日本 その他（ ）

(6) あなたにとっての母語（○をしてください） 日本語 その他（ ）

(7) 英語の学習年数（英語を学び始めてから今日まで）をお答えください。
（ ）年（ ）ヶ月

(8) 英語の学習場所は、中学校や高等学校などの学校の英語の授業でしたか。（○をしてください）
いいえ どちらかといえばいいえ ややはい はい

(9) 前の質問 (8) で、「いいえ」と「どちらえといえばいいえ」に○をした人にききます。主な英語の学習場所はどこですか。

(10) 海外経験はありますか。（○をしてください） ない ある
（あると答えた方へ 簡単に、期間と場所、目的[旅行、留学、ホームステイなど]を記入してください。例、「アメリカ・4 日間・観光旅行・小 4 のとき」など。）

(11) 英語に関する資格について教えてください。
（英語検定、TOEIC、TOEFL など。得点や級と受験時期を差し支えない程度に。例、「英検 4 級・中 1 のとき」など）

APPENDIX C

FOLLOW-UP QUESTIONNAIRE

事後アンケート 2

氏名（　　　　　　　　　　　　　　　　）　　日付（　　　　　　　　）

次の文法規則について、①現在、知っているか、②知っているならば、いつ知ったか、について教えてください。①と②について、それぞれ該当項目に○をしてください。

A. 英語では、2 つ以上の形容詞（*two big round white wooden* balls のイタリック体の単語）を並べるときの順序がほぼ決まっている。

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（　）

B. 英語では形容詞の順序は決まっているが、日本語では形容詞の語順は決まっておらず柔軟である。例えば、英語では "a big white ball" と言えるが、"a white big ball" と言えない。日本語では、「大きくて白いボール」も「白くて大きいボール」も言える。

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（　）

C. 英語の形容詞の語順は、「形」に関する形容詞（big、tall、small、long など）と「色」に関する形容詞（white、brown、green など）では、「形」+「色」の順番（a *long green* pen）となる。

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（　）

D. **英語では、動詞の目的語を主語にし、be 動詞を挿入し、動詞を過去分詞形にすることによって、「～は・・・される」という受動態（受け身形）を表現することができる。**

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（ ）

E. **英語の受動態では、動詞の目的語しか主語にできないが、日本語の受動態では、動詞の目的語でなくても、被害を受ける主体を主語にすることができる。例えば、Mike uses this computer. という英文では、use の目的語の this computer しか、受動態の受け身形の主語にならない（例、This computer is used by Mike.）。一方、日本語では、「クミはマイクに彼女のコンピュータを使われた。そのため、彼女は困った。」というように、被害を受けた主体（クミ）を主語にすることができる。**

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（ ）

F. **「クミはマイクに彼女のコンピュータを使われた」のように被害を表すときには、Kumi had her computer used by Mike. のように、「have +目的語+過去分詞形」の形で表現する。**

① 1. 知っている 2. 言われれば聞いたことがあると思う
3. 知らなかった（今、初めて知った）

② 1. 中学校のとき 2. 高校のとき 3. 予備校のとき
4. 大学の授業で 5. 今回の実験の時 6. その他（ ）

今回の実験の最中（1 回目から 3 回目までの間）に、家庭で、あるいは大学の授業などで、上記の A から F までの規則について学んだり、調べたりしましたか。

1. はい 2. いいえ
3. いくつかについては「はい」
（その文法規則を記号で書いてください ）

APPENDIX D

META-LANGUAGE TEST

氏名　　　　　　　　　　　　日付

次の英文には、一か所文法的な誤りがあります。そこの部分に下線を引き、どのように間違っているのかを説明しなさい。文法用語を使っても使わなくても良いです。もし誤りが見つからない場合は、「正しいと思います」と書いてください。

1	I found a blue large book in the library.	
2	Taro was read his love letter by his mother.	
3	Mary speaked Chinese with that man.	

APPENDIX E

EXIT QUESTIONNAIRE

アンケート（事後用）

あなたについて教えてください。

(1) 日付 2012 年（ ）月（ ）日

(2) 氏名 （ ）

(3) 中学校教員免許（外国語・英語）を取得予定ですか。（○をしてください）

一種・主免許　一種・副免許　二種・副免許　予定なし

(4) 研究に協力する時点で次の授業を受講していましたか。あるいはしていますか。

(a) 田中江扶先生の英語学関係の授業（総合演習などは含みません）はい いいえ

(b) 高橋渉先生の英語学関係の授業（コンピュータ利用教育などは含みません）
はい　いいえ

(c) コリーン・ダルトン先生の授業　はい　いいえ

(d) その他の英語学関係の授業　はい（ ）　いいえ

(5) 本研究が対象としている内容について、あなたはどう思いましたか。<u>タスクやテストをしているときに感じていたこと</u>を答えてください。1 ～ 5 に当てはまる数字に○をしてください。

まったくそう思わなかった	どちらかといえばそう思わなかった	どちらともいえない	どちらかといえばそう思った	まったくそう思った
1	2	3	4	5

(a) 研究の目的は、複数形の使用を調べることであると思った。
1-----2-----3-----4-----5

(b) 研究の目的は、どれだけたくさん話せるかを調べることであると思った。
1-----2-----3-----4-----5

(c) 研究の目的は、過去形の使用を調べることであると思った。
1-----2-----3-----4-----5

(d) 研究の目的は、単語力を調べることであると思った。 1-----2-----3-----4-----5

(e) 研究の目的は、形容詞の使用を調べることであると思った。
1-----2-----3-----4-----5

(f) 研究の目的は、英語で言えない表現があったときに、どうやって言い換えるかを調べることであると思った。 1-----2-----3-----4-----5

(g) 研究の目的は、受動態（受身形）を調べることであると思った。
1-----2-----3-----4-----5

(h) 研究の目的は、どれぐらい流暢に英語が話せるかを調べることであると思った。

1-----2-----3-----4-----5

(6) 研究に協力してみた感想を書いてください。

(7) 機会があったらまた協力していただけますか。 はい いいえ

(8) データを使ってもよいですか。 はい いいえ

APPENDIX F

PARTICIPANTS' INFORMATION (ADJECTIVE ORDERING)

	Group				
	POS	REC	REC+	SAL	Statistical analysis
n	20	20	16	19	
REL					
Teacher	17	15	13	16	Fisher's exact test
Researcher	3	5	3	3	$p = .880$
EXP					
Yes	9	10	7	9	Fisher's exact test
No	11	10	9	10	$p = .989$
PREV					
Yes	6	4	4	4	Fisher's exact test
Maybe Yes	9	13	6	11	$p = .873$
No	4	2	5	3	
No response	1	1	1	1	
MLT					
Wrong answer	11	11	9	9	Fisher's exact test
Without metalanguage	7	7	4	7	$p = .971$
With metalanguage	2	2	3	3	
EXIT					
M	3.30	3.10	3.63	3.79	One-way ANOVA
SD	0.98	1.29	1.09	1.13	$F(3, 71) = 1.45$, $p = .235$

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; REL = relationship with the researcher; Teacher = teacher-student relationship; Researcher = researcher-participant relationship; EXP = experience abroad; PREV = previous knowledge about the ungrammaticality of the color-size-noun sequence in English; MLT = the scores of the meta-language tests that were administered in the last session; Without metalanguage = correct explanation without using metalanguage; With metalanguage = correct explanation with metalanguage; EXIT = participant's awareness about the purpose of the study (The item in the exit questionnaire was, "Did you think that the purpose of this study was to examine the knowledge about adjectives?").

APPENDIX G

PARTICIPANTS' INFORMATION (INDIRECT PASSIVES)

	Group				
	POS	REC	REC+	SAL	Statistical analysis
N	21	22	27	20	
REL					
Teacher	18	16	19	16	Fisher's exact test
Researcher	3	6	8	4	$p = .624$
EXP					
Yes	10	12	11	10	Fisher's exact test
No	11	10	16	10	$p = .810$
PREV					
Yes	7	9	8	7	Fisher's exact test
Maybe Yes	4	7	14	8	$p = .455$
No	9	5	4	4	
No response	1	1	1	1	
MLT					
Wrong answer	14	19	18	12	Fisher's exact test
Without metalanguage	2	1	2	2	$p = .615$
With metalanguage	5	2	7	6	
EXIT					
M	3.95	3.77	4.22	4.04	One-way ANOVA
SD	1.12	0.75	0.75	0.89	$F(3, 86) = 1.35$, $p = .264$

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; REL = relationship with the researcher; Teacher = teacher-student relationship; Researcher = researcher-participant relationship; EXP = experience abroad; PREV = previous knowledge about the ungrammaticality of indirect passives in English; MLT = the scores of the meta-language tests that were administered in the last session; Without metalanguage = correct explanation without using metalanguage; With metalanguage = correct explanation with metalanguage; EXIT = participant's awareness about the purpose of the study (The item in the exit questionnaire was, "Did you think that the purpose of this study was to examine the knowledge about passivization?").

APPENDIX H

THE SPOT-THE-DIFFERENCE TASK

Instructions: Picture 1 is the original, and Picture 2 is a copy of that picture. What is different from the original picture? Find the differences and describe the differences in the copy picture in English. When you report the differences, please start with a phrase, "In picture 1, there is/are ---. But in picture 2, there is/are ---."

指示文。オリジナル（ピクチャー1）の絵と変わっているものは何ですか。変わっているところを見つけて、英語で説明してください。できるだけたくさんの違いを見つけてください。見つけた違いは、全部説明してください。英語で説明するとき、次の表現をつかってください。

In picture 1, there (is/are) ～. But in picture 2, there (is/are) ～.

Expected phrases:

Picture 1: The Original Picture	Picture 2: The Copy Picture
a short blue boat	a long red boat
a large yellow computer	a small blue computer
a low blue chair	a high gray chair
a small gray radio	a large orange radio

Pictures:

Picture 1 (Original)	Picture 2 (Copy)
	

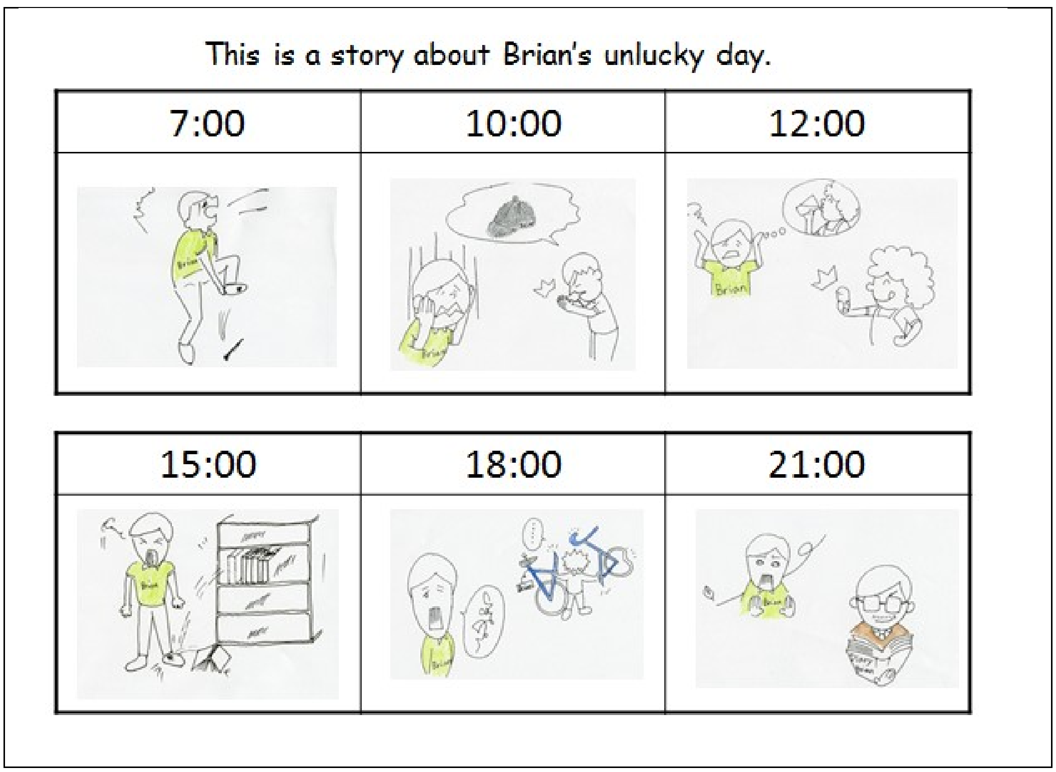
APPENDIX I

THE PICTURE-CUED STORY MAKING TASK

Instructions: This is a task in which you will make a story in English, looking at six pictures. The six pictures depict Brian’s unlucky day. Now you will hear the brief explanation of each picture in Japanese. (Participants listen to the Japanese explanations and look at the pictures). Now you will make a story about him in English. You can create your own story or change the explanations. Make a story covering the six pictures. Start with the following sentence: This is a story about Brian’s unlucky day.

指示文。これは、6枚の絵を見て、英語で物語を作るタスクです。6枚の絵は、ブラアンの、ついていない一日 (Brian’s unlucky day) を表したものです。これから、それぞれの絵について簡単に説明をします。（絵の説明を聞きながら絵を見る）それでは、物語作りタスクを始めます。6枚の順番をシャッフルします。ブライアンについて物語を英語で作りなさい。自分で話を創作してかまいません。つまり、追加したり、変更してもかまいません。それぞれの絵がつながるように話を作って下さい。次の英文から開始してください。This is a story about Brian’s unlucky day.

Pictures (from Form A)



Expected (incorrect) sentences (Form A):

Picture	Expected sentences
Picture 1	distractor
Picture 2	*At 10:00 am, he was *lost* his nice cap by his brother.
Picture 3	*At noon, he was *drunk* his juice by his mother.
Picture 4	distractor
Picture 5	*At 6:00 pm, he was *broken* his nice bike by his friend.
Picture 6	*At 9:00 pm, he was *read* his diary by his father.

Explanation of each picture in Japanese (Form A):

Picture	Explanation of each picture
Picture 1	ブライアンが歩いているときに、大きなくぎを踏みつけてしまいました。
Picture 2	ブライアンは、お兄さんに彼の素敵な帽子をなくされました。その帽子は、黒くて、お洒落でした。ブライアンは、その帽子を愛用していました。
Picture 3	ブライアンは牛乳を飲むのが好きでした。しかし、ブライアンは、お母さんに彼の牛乳を飲まれました。
Picture 4	ブライアンの足の上に、本が落ちてきました。その本はとても大きく重い本でした。とても痛く感じました。
Picture 5	ブライアンは、友だちに彼の素敵な自転車を壊されました。その自転車は、青くて高価なものでした。
Picture 6	ブライアンは、お父さんに彼の日記を読まれました。

APPENDIX J

THE ITEMS FOR ELICITED IMITATION TEST

Instructions: You will listen to an English sentence one time for each question. Choose the picture that the sentence describes. Then, repeat the sentence in correct English.

指示文。各問題は、一度しか英文が流れません。英文を聞いたら、**すぐに、**(1) 英文を選んで（記号をいってください）、(2) 繰り返してください。文法的に正しくない英文は、正しく直して繰り返してください。

Test Sentences
Adjective Ordering
*There is a *gray large picture* on the wall.
*Koji is opening the *red small door* now.
*I bought a *yellow long bed* yesterday.
*Lisa likes the *blue high mountain* very much.
*There is an *orange short bus* on the street.
Mary received a *large yellow card* last night.
There is a *small gray box* near the ball.
There is a *long blue river* in the town.
My cat is sleeping on the *high orange table* now.
Jim has a *big red cat* at home.
Indirect Passives
*David was smoked a cigarette by a lady.
*Taro was killed his dog.
*Bob was eaten his cookies by his sister.
*Hiroshi was broken his toy car.
*Mike was stolen his nice pen.
Brian was taught how to use a computer at the school.
David was bought a new coat last Sunday.
Ken was given a new bike as a birthday present.
Yumi was shown the picture at first.
Jane was told the story by Mike.
Distractors
*The boy feeded the cat this morning.
*Last night I beated my father at shogi.
*The stone striked the girl on her head yesterday.

(*Table continues*)

(*Table continued*)

Test Sentences
*Yumi eated two hamburgers last night.
*Koji teached Japanese to Brian yesterday.
Yesterday my brother found a new restaurant.
An old man stood alone near the tree last night.
Mike read a long mystery last week.
Jim gave a nice present to his mother yesterday.
Cathy bought a nice hat this evening.

APPENDIX K

PICTURES FOR THE ELICITED IMITATION TEST

Picture 1 (correct)	Picture 2	Picture 3
Adjective Ordering		
a large gray picture	a large blue picture	a small gray picture
a small red door	a small yellow door	a large red door
a long yellow bed	a long blue bed	a short yellow bed
a high blue mountain	a high black mountain	a low blue mountain
a short orange bus	a short green bus	a long orange bus
a large yellow card	a large pink card	a small yellow card
a small gray box	a small white box	a large gray box
a long blue river	a long brown river	a short blue river
a high orange table	a high red table	a low orange table
a big red cat	a big blue cat	a small red cat
Indirect Passives		
David was upset because a lady smoked near him.	David was upset because a man smoked near him.	Jane was upset because a lady smoked near her.
His dog was killed.	His cat was killed.	Taro and his dog were alive and happy.
Bob was upset because his sister ate his cookies.	Bob was upset because his sister ate his cake.	His sister was upset because Bob ate her cookies.
Hiroshi was upset because his toy car was broken.	Hiroshi was upset because his toy bike was broken.	Hiroshi was playing with his toy car.
Mike was upset because his nice pen was stolen.	Mike was upset because his wallet was stolen.	Mike was writing with a nice pen.
Brian was taught how to use a computer at the school.	Brian was taught how to use a laundry machine.	Brian was teaching how to use a computer.
David was bought a new coat last Sunday.	David was bought a new hat.	David bought a new coat.
Ken was given a new bike as a birthday present.	Ken was given a new bag.	Ken was given an old bike.
Yumi was shown the picture at first.	Yumi was shown the book.	Yumi showed the picture.

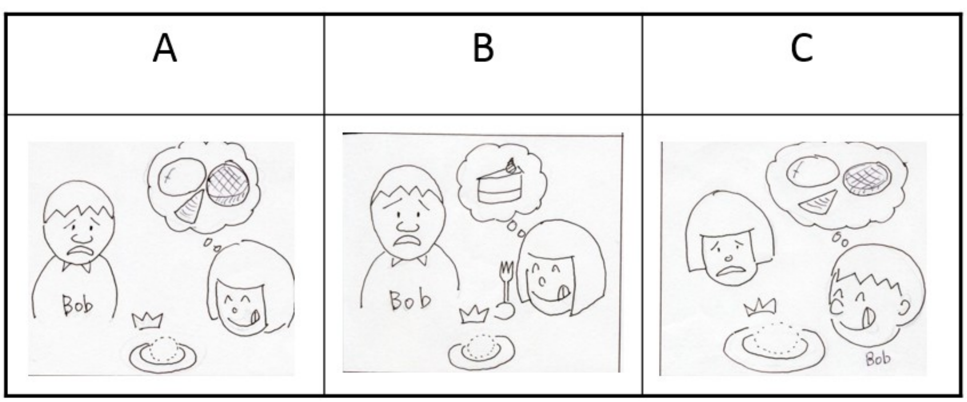
(*Table continues*)

(*Table continued*)

Picture 1 (correct)	Picture 2	Picture 3
Jane was told the story by Mike.	Jane told the story.	Jane was reading a book.
Distractors		
The boy fed the cat this morning.	The boy fed the horse.	The boy fed the dog.
Last night I beat my father at shogi.	I beat my father at chess.	My father beat me at shogi.
The stone struck the girl on her head yesterday.	The ball struck the girl on her head.	The stone struck the boy on the head.
Yumi ate two hamburgers last night.	Yumi ate one hamburger.	Ken ate two hamburgers.
Koji taught Japanese to Brian.	Koji taught Chinese to Brian.	Brian taught Japanese to Koji.
Yesterday my brother found a new restaurant.	My brother found a new park.	My sister found a new restaurant.
An old man stood alone near the tree last night.	A young man stood alone.	Two men stood by the tree.
Mike read a long mystery last week.	Mike wrote a story.	Mike bought a book.
Jim gave a nice present to his mother yesterday.	Jim gave a present to his father.	His mother gave a present to Jim.
Cathy bought a nice hat this evening.	Cathy bought a bag.	Cathy made a hat.

Note. The order of the pictures for each item was randomized.

Sample Picture (**Bob was eaten his cookies by his sister.*)**



APPENDIX L

THE ITEMS FOR GRAMMATICALITY JUDGMENT TEST

Instructions: Some of the following sentences are grammatically correct, e.g., in terms of word orders and word forms; others are not. Judge whether each sentence is grammatically correct or not, according to your knowledge about English. Indicate the number from 1 (*incorrect*) to 4 (*correct*). Choose *Correct* or *Somewhat correct* when you think, "I would say the meaning in the same way as the sentence." Choose *Incorrect* or *Somewhat incorrect* when you think, "I would not say like the sentence and express the meaning in a different way."

指示文。次の英文の中には、文法的に正しい文と、語順や語形など、文法的に正しくない文が含まれています。それぞれ、文法的に正しいか、文法的に正しくないか、あなた自身の英語の知識に照らし合わせて、判断してください。「正しくない」から「正しい」までの判断を、数字で解答欄に記入してください。もし「自分も示された英文のように言うだろうな」と思うときは、「正しい」か「やや正しい」を選んでください。もし「自分なら別の言い方をして、示された英文のようには言わない」と感じるときは、「正しくない」か「あまり正しくない」を選んでください。

Test Sentences
Adjective Ordering
*I found a *blue large book* in the library.
*I see a *yellow small dog* every morning.
*Yumi is drawing an *orange long line* on the paper.
*There are *red high walls* around the house.
*There is a *gray tall glass* on the desk.
My father has a large red building in the city.
There is a small orange ball in the box.
Ken is walking along the long gray street.
I saw a high yellow ship yesterday.
There is a little blue phone on the table.
Indirect Passives
*Taro was *read* his love letter by his mother.
*Brian was *kicked* his leg by the man in the train.
*Bob was *scolded* his daughter by a man.
*Kumi was *lost* her necklace by her sister.
*Jane was *broken* her new computer yesterday.
Yumi was taught his language by Reymond.
David was bought a new mobile phone last Sunday.

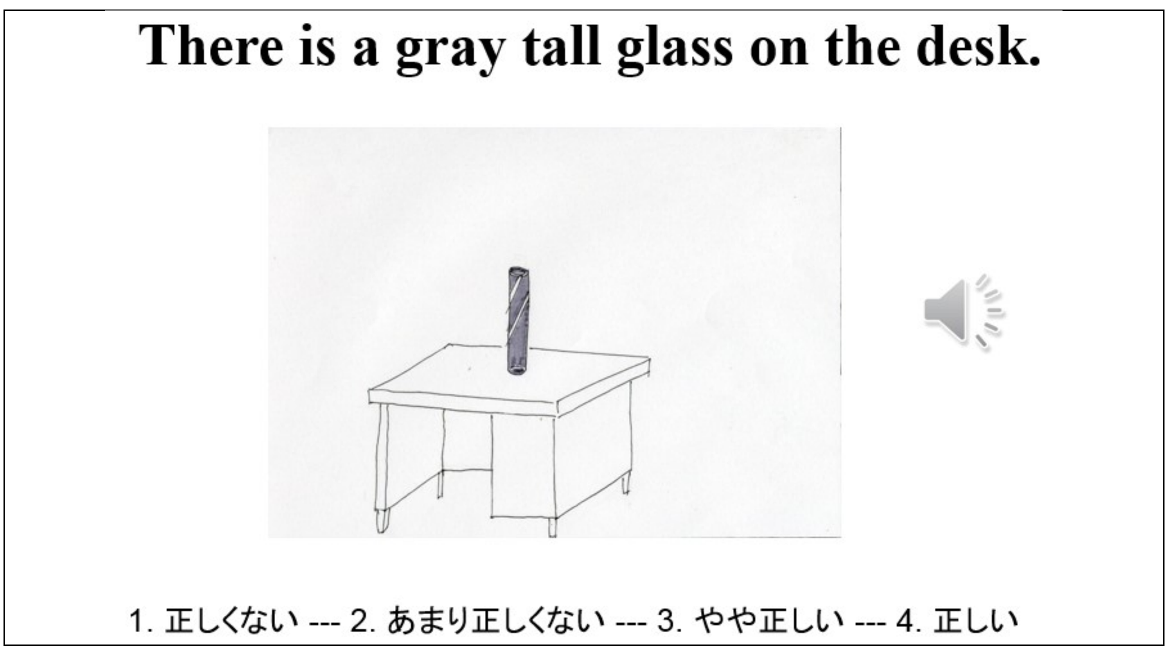
(*Table continues*)

(*Table continued*)

Test Sentences
Hideki was given his first watch as a Christmas present.
Yoko was shown his ID card at first.
Michi was told his car was stolen.
Distractors
*The book falled off my desk.
*In the lesson, the teacher hanged a map on the blackboard.
*Mary speaked Chinese with that man.
*Jane eated two cakes last night.
*Brian teached American history to Koji.
Yesterday my brother took a picture of the station.
A girl stood alone in the park yesterday.
Jun saw a famous singer in the city last Sunday.
Wataru went to Tokyo last weekend.
My father made a nice dog house last year.

Note. Each sentence is attached with a picture depicting the meaning of the sentence.

An Example of Slides for the Grammaticality Judgment Test



APPENDIX M

THE OBJECT PLACEMENT TASK

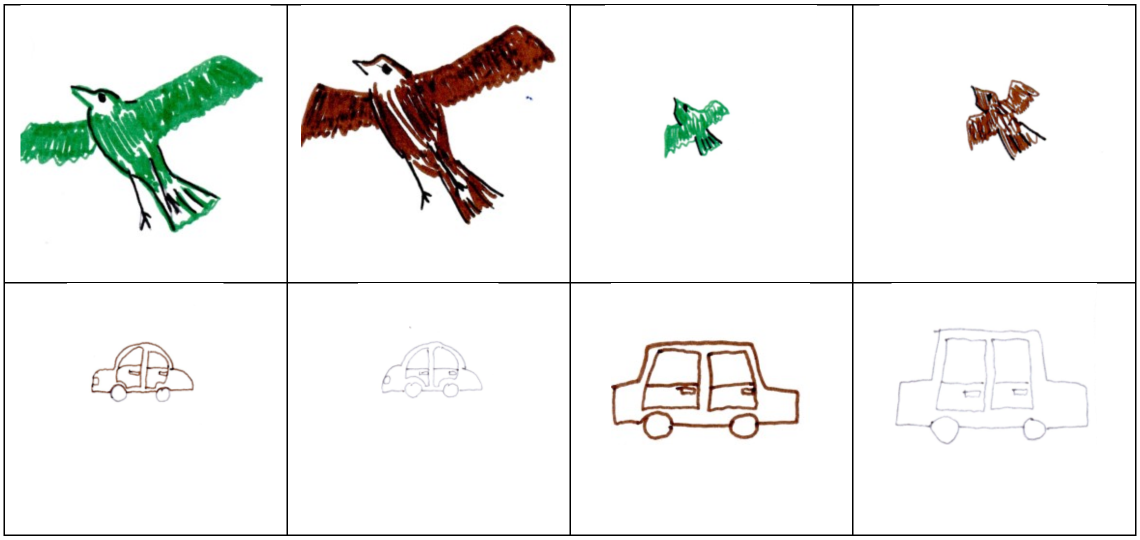
Instruction: You will have 20 small picture cards and 5 name tags. You will hear the instructions in English. Please choose appropriate items and place them on the small board illustrated with an outdoor scene.

指示文。20枚の小さい絵と5枚の名前があります。英語で指示が与えられます。適切な絵や名前を選び、パノラマの上に置いてください。

Instruction
Adjective Ordering
Please put a *big green bird* under the clouds.
Please put a *little brown car* in front of the house.
Please put a *short white tower* in front of the lake.
Please put a *tall black tree* by the street.
Please put a *tall pink cup* on the table.
Indirect Passives
Ken was upset because his wallet was stolen at the station.
Taro was upset because his sister ate his cake.
Yumi was upset because her nice glass was broken.
Keiko was upset because a young lady scolded her dog.
Jane was upset because her brother lost her watch.

Picture Cards				
a big green bird	a big brown bird	a small green bird	a small brown bird	
a little brown car	a little white car	a big brown car	a big white car	
a short white tower	a short pink tower	a tall white tower	a tall pink tower	
a tall black tree	a tall green tree	a short black tree	a tall green tree	
a tall pink cup	a tall brown cup	a short pink cup	a short brown cup	
Name Tags				
Ken	Taro	Keiko	Yumi	Jane

Examples of Picture Cards



Small Board

APPENDIX N

PICTURES FOR THE PICTURE IDENTIFICATION TASK

Instructions (output activity): You and I have the same sets of pictures. Please describe one of the four pictures in English. I (the researcher) will identify that picture.

指示文（アウトプット活動）。あなたと私は、同じ絵カードを持っています。4枚の絵の中から1つ選んで、英語で描写してください。私は、あなたが説明する絵を探します。

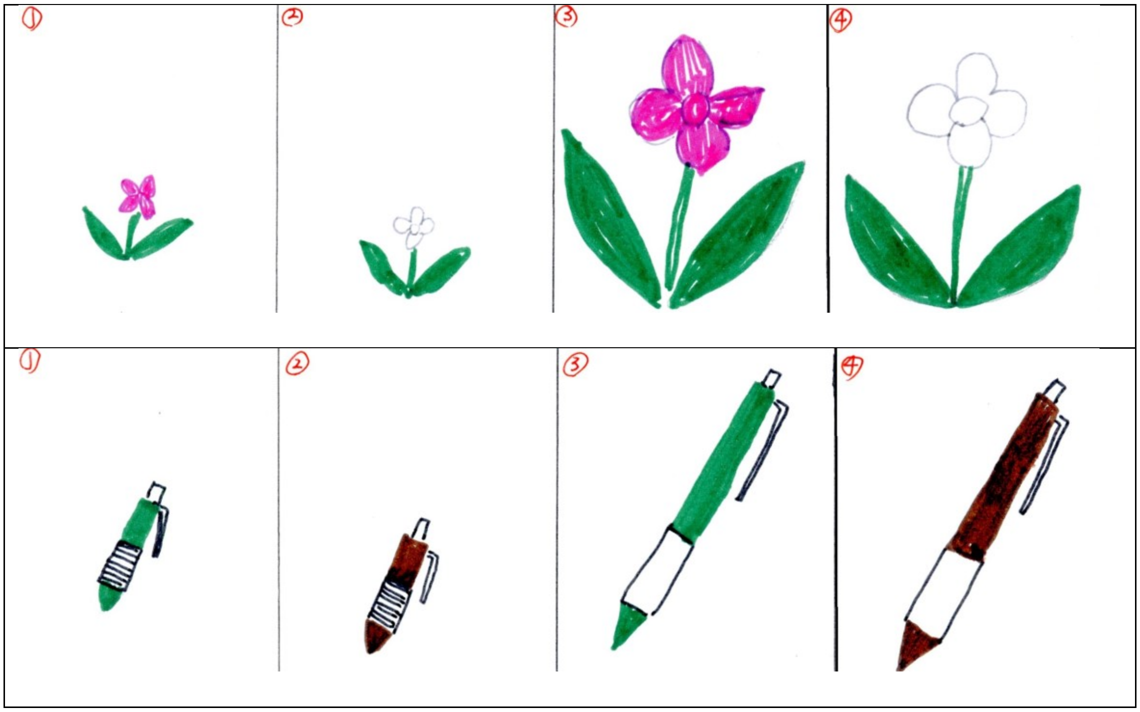
Instructions (as an input activity): You and I have the same sets of pictures. You will hear a description of one picture in English. Please identify that picture.

指示文（インプット活動）。あなたと私は、同じ絵カードを持っています。英語で1枚の絵について説明を聞きます。その絵を探してください。

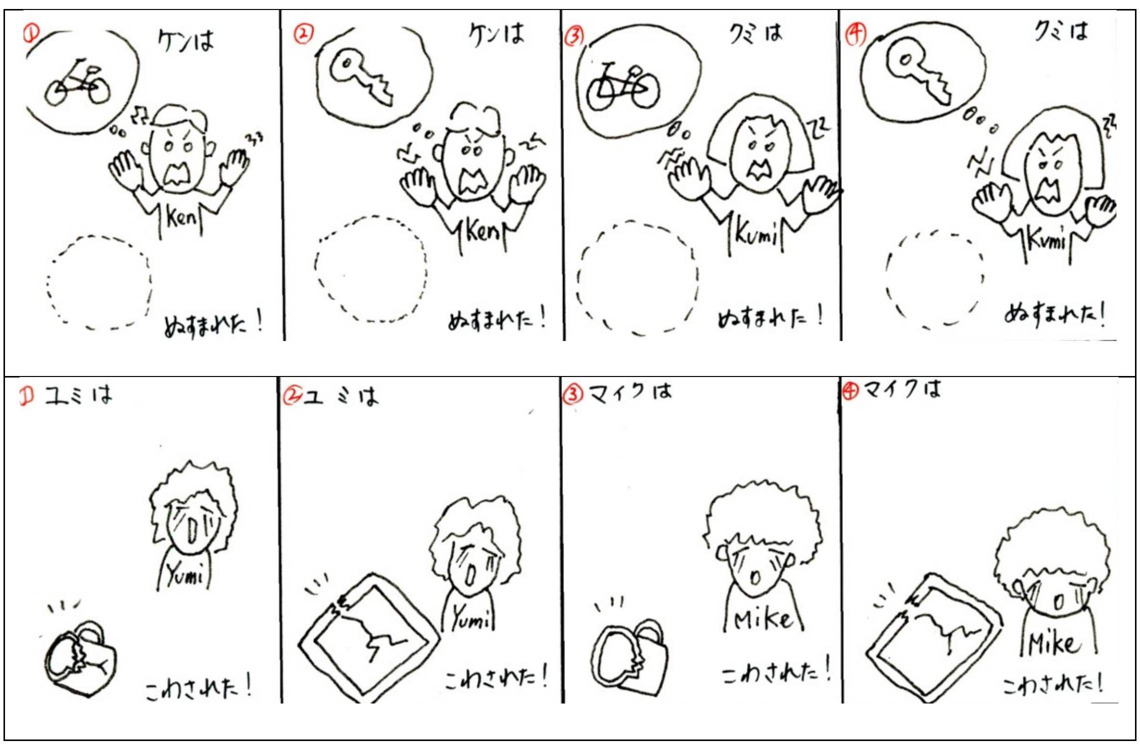
Picture 1 (Target)	Picture 2	Picture 3	Picture 4
Adjective Ordering			
a big white bag	a big black bag	a small white bag	a small black bag
a little pink flower	a little white flower	a big pink flower	a big white flower
a short green pen	a short brown pen	a long green pen	a long brown pen
a short black T-shirt	a short pink T-shirt	a long black T-shirt	a long pink T-shirt
a tall brown house	a tall green house	a small brown house	a small green house
Indirect Passives (The following were expected errors.)			
Ken was stolen his bike.	Ken was stolen his key.	Kumi was stolen her bike.	Kumi was stolen her key.
Taro was eaten his ice cream by Kumi.	Taro was eaten his steak by Kumi.	Kumi was eaten her ice cream by Taro.	Kumi was eaten her steak by Taro.
Yumi was broken her nice cup.	Yumi was broken her nice picture.	Mike was broken his nice cup.	Mike was broken his nice picture.
Keiko was scolded her son by a man.	Keiko was scolded her son by a woman.	Jim was scolded his daughter by a man.	Jim was scolded his daughter by a woman.
Jane was lost her book by Peter.	Jane was lost her glass by Peter.	Peter was lost his book by Jane.	Peter was lost his glass by Jane.

Note. The order of the four pictures for each item was randomized.

Example of Pictures on Cards (Adjective Ordering)



Example of Pictures on Cards (Indirect Passives)



APPENDIX O

EFFECT SIZES (R) OF THE GROUP COMPARISONS

ON THE THREE ADJECTIVE-ORDERING TESTS

	Pre-Post	Pre-Delayed	Post-Delayed
OPT			
POS vs REC	.21	.01	.09
POS vs REC+	.10	.22	.19
POS vs SAL	.29	.22	.04
REC vs REC+	.09	.27	.29
REC vs SAL	.14	.24	.14
REC+ vs SAL	.22	.02	.11
EIT			
POS vs REC	.17	.06	.12
POS vs REC+	.20	.09	.13
POS vs SAL	.01	.04	.08
REC vs REC+	.03	.06	.03
REC vs SAL	.16	.05	.22
REC+ vs SAL	.18	.04	.26
GJT			
POS vs REC	.32	.33	.16
POS vs REC+	.19	.32	.07
POS vs SAL	.12	.02	.16
REC vs REC+	.13	.13	.21
REC vs SAL	.18	.37	.06
REC+ vs SAL	.06	.36	.21

Note. OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period. The shadowed cells indicate a medium effect size; The criteria were .10 for a small effect, .30 for a medium effect, and .50 for a large effect.

APPENDIX P

CONTINGENCY TABLES FOR INDIVIDUAL ANALYSES OF PERFORMANCE ON THE THREE ADJECTIVE-ORDERING TESTS

		Contingency Table A		Contingency Table B		Contingency Table C	
Group	*n*	L1 transfer	Other	P+D	Other	P+D or D	Other
OPT							
POS	20	3	17	0	3	1	2
REC	20	2	18	1	1	2	0
REC+	16	1	15	0	1	0	1
SAL	19	4	15	0	4	1	3
		p = .656		*p* = .300		*p* = .429	
EIT							
POS	20	13	7	0	13	3	10
REC	20	11	9	3	8	3	8
REC+	16	12	4	3	9	4	8
SAL	19	9	10	2	7	2	7
		p = .384		*p* = .188		*p* = .937	
GJT							
POS	20	20	0	0	20	1	19
REC	20	13	7	0	13	2	11
REC+	16	16	0	1	15	3	13
SAL	19	15	4	1	14	1	14
		p = .050		*p* = .563		*p* = .561	

Note. OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; L1 transfer = the number of the participants who produced or accepted both the incorrect adjective order (color-size) and the correct adjective order (size-color); P+D = the number of the participants who performed better on the posttest and maintained the improvement on the delayed posttest; P+D or D refers to the number of the participants who performed better on the posttest and maintained the improvement on the delayed posttest (P+D) or who performed better on the delayed posttest (D).

APPENDIX Q

EFFECT SIZES (R) OF THE GROUP COMPARISONS

ON THE THREE PASSIVIZATION TESTS

	Pre-Post	Pre-Delayed	Post-Delayed
OPT			
POS vs REC	.03	.15	.08
POS vs REC+	.17	.19	.26
POS vs SAL	.09	.16	.02
REC vs REC+	.19	.04	.19
REC vs SAL	.05	.01	.05
REC+ vs SAL	.26	.04	.22
EIT			
POS vs REC	.41	.28	.20
POS vs REC+	.15	.14	.09
POS vs SAL	.28	.36	.08
REC vs REC+	.24	.11	.13
REC vs SAL	.16	.09	.32
REC+ vs SAL	.12	.19	.18
GJT			
POS vs REC	.05	.01	.04
POS vs REC+	.07	.28	.25
POS vs SAL	.17	.30	.24
REC vs REC+	.14	.26	.17
REC vs SAL	.20	.28	.19
REC+ vs SAL	.06	.04	.03

Note. OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period. The criteria were .10 for a small effect, .30 for a medium effect, and .50 for a large effect.

APPENDIX R

CONTINGENCY TABLES FOR INDIVIDUAL ANALYSES OF PERFORMANCE ON THE THREE PASSIVIZATION TESTS

		Contingency Table A		Contingency Table B		Contingency Table C	
Group	*n*	L1 transfer	Other	P+D	Other	P+D or D	Other
OPT							
POS	21	3	18	0	3	0	3
REC	22	2	20	0	2	1	1
REC+	27	7	20	0	7	0	7
SAL	20	2	18	1	1	1	1
		p = .394		p = .286		p = .099	
EIT							
POS	21	12	9	0	12	0	12
REC	22	14	8	0	14	0	14
REC+	27	17	10	1	16	1	16
SAL	20	13	7	0	13	0	13
		p = .972		p = 1.000		p = 1.000	
GJT							
POS	21	5	16	0	5	0	5
REC	22	8	14	0	8	0	8
REC+	27	11	16	2	9	3	8
SAL	20	7	13	1	6	2	5
		p = .679		p = .667		p = .292	

Note. OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; L1 transfer = the number of the participants who produced or accepted both the incorrect adjective order (color-size) and the correct adjective order (size-color); P+D = the number of the participants who performed better on the posttest and maintained the improvement on the delayed posttest; P+D or D refers to the number of the participants who performed better on the posttest and maintained the improvement on the delayed posttest (P+D) or who performed better on the delayed posttest (D).

APPENDIX S

INDIVIDUAL FACTORS (ADJECTIVE ORDERING)

Participant	OPT	EIT	GJT	CELT	REC	PREV	MLT	EXIT
POS								
1078			Delayed	30		MY	1	4
1028	Delayed	Delayed		42		NA	1	5
1013		Delayed		48		Y	1	3
1092		Delayed		38		N	0	3
1023		Short	Short	22		N	0	4
1096	Short					MY	0	2
1009				38		MY	0	4
1029				47		Y	2	4
1043				55		Y	0	4
1045				54		MY	1	4
1046				66		MY	2	3
1050				28		N	1	2
1066				49		MY	1	4
1071				59		Y	0	4
1076				38		MY	0	3
1086				31		MY	0	3
1098				26		N	0	3
1103				27		Y	0	2
1109						Y	0	4
1111				37		MY	1	1
REC								
1034		Retained	Short	59	1	MY	0	1
1007		Retained	Delayed	46	1	Y	1	4
1070		Retained		62	3	N	1	4
1083	Retained			40	5	MY	0	1
1100			Delayed	40	2	MY	1	3
1093	Delayed			10	1	MY	0	1
1026		Short	Short	49	1	MY	1	3
1002				47	1	Y	0	4
1011				39	1	NA	1	3
1018				41	1	MY	0	2
1022				51	1	MY	1	4
1035				43	1	MY	0	4
1044				52	3	MY	2	5
1065				40	2	N	1	4
1073				48	1	Y	2	4
1082				36	2	MY	0	2

(*Table continues*)

(*Table continued*)

Participant	OPT	EIT	GJT	CELT	REC	PREV	MLT	EXIT
1088				32	4	MY	0	1
1095				47	2	MY	0	4
1099				26	2	Y	0	4
1118				23	2	MY	0	4
REC+								
1112		Retained	Retained	42	1	Y	2	5
1017		Retained		55	3	MY	2	4
1114		Retained		67	1	N	0	2
1064		Delayed	Delayed	43	3	Y	1	4
1119			Delayed	23	5	MY	0	2
1019		Short	Short	50	1	N	1	5
1069			Short	45	1	N	0	4
1115		Short		31	1	MY	0	4
1012				53	2	Y	0	4
1021				46	1	NA	0	4
1051				40	2	MY	2	5
1072				49	2	Y	1	4
1087				39	2	MY	0	2
1090				25	2	N	1	2
1116				36	2	MY	0	4
1121				37	4	N	0	3
SAL								
1005			Retained	35		Y	1	5
1006		Retained	Short	41		MY	0	5
1081		Retained		46		MY	0	5
1107	Delayed			32		MY	0	3
1074	Short			48		MY	0	5
1001				59		Y	1	3
1010				32		NA	2	4
1014				51		MY	1	2
1020				52		Y	1	4
1024				50		N	2	2
1031				57		Y	1	5
1040				51		MY	1	4
1048				52		MY	0	4
1068				34		MY	1	2
1080				57		MY	0	5
1089				55		N	2	5
1102				19		MY	0	3
1108				28		MY	0	3
1110				26		N	0	3

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; CELT

= the results of the structure section of the Comprehensive English Language Test for Learners of English (CELT Form A, Harris & Palmer, 1986); REC = the number of recasts provided in the treatment task; PREV = previous knowledge; MLT = the scores of the meta-language tests; EXIT = participant's awareness about the purpose of the study. Retained (that is, retained effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the posttest. Y = yes; N = no; MY = maybe yes.

APPENDIX T

INDIVIDUAL FACTORS (INDIRECT PASSIVES)

Participant	OPT	EIT	GJT	CELT	REC	PREV	MLT	EXIT
POS								
1043		Short		55		Y	2	5
1009				38		N	0	4
1013				48		Y	0	3
1023				22		N	0	3
1028				42			0	5
1029				47		MY	1	2
1045				54		Y	0	5
1046				66		MY	2	5
1050				28		N	2	4
1066				49		N	0	5
1071				59		N	0	5
1076				38		Y	0	3
1078				30		MY	0	4
1086				31		Y	1	5
1092				38		N	0	4
1096						Y	2	2
1098				26		N	0	4
1103				27		N	0	2
1106				56		Y	2	5
1109						MY	0	5
1111				37		N	0	3
REC								
1035	Delayed			43	4	Y	0	4
1002				47	4	MY	0	4
1004				43	5	MY	0	4
1007				46	4	N	0	3
1011				39	3		0	3
1022				51	1	N	0	4
1026				49	2	MY	0	3
1038				38	1	Y	2	4
1042				48	2	Y	0	4
1065				40	4	MY	0	4
1073				48	2	Y	2	4
1079				41	4	N	0	4
1082				36	2	MY	0	5
1083				40	4	N	0	2
1085				31	4	N	0	4

(*Table continues*)

(*Table continued*)

Participant	OPT	EIT	GJT	CELT	REC	PREV	MLT	EXIT
1088				32	4	MY	0	4
1093				10	4	Y	0	4
1095				47	3	MY	0	5
1099				26	4	Y	0	2
1100				40	4	Y	0	4
1117				50	1	Y	1	4
1118				23	1	Y	0	4
REC+								
1051		Retained	Retained	40	2	Y	2	5
1112			Retained	42	2	MY	1	5
1075			Delayed	59	2	MY	2	4
1049	Delayed			31	4	N	0	5
1064	Delayed			43	4	MY	0	4
1091	Delayed			38	4	Y	0	4
1119	Delayed			23	2	Y	2	5
1003				47	3	MY	0	4
1012				53	4	MY	2	2
1017				55	4	N	2	4
1030				52	3		0	3
1036				44	5	MY	0	4
1037				39	2	MY	0	4
1069				45	1	Y	1	5
1072				49	1	Y	2	4
1077				43	1	MY	0	4
1084				51	5	Y	2	5
1087				39	4	MY	0	4
1090				25	4	N	0	5
1094				54	4	Y	0	4
1097				35	5	N	0	4
1104				31	3	MY	0	3
1105				38	1	Y	0	4
1115				31	4	MY	0	5
1116				36	3	MY	0	5
1120				47	1	MY	0	4
1121				37	4	MY	0	5
SAL								
1089			Retained	55		MY	2	5
1020	Retained			52		MY	2	4
1040			Delayed	51		Y	2	5
1001				59		Y	2	5
1005				35		N	0	5
1006				41		Y	0	4

(*Table continues*)

(*Table continued*)

Participant	OPT	EIT	GJT	CELT	REC	PREV	MLT	EXIT
1010				32			0	4
1014				51		N	0	4
1024				50		Y	0	4
1031				57		Y	2	5
1041				50		N	0	2
1048				52		MY	0	2
1068				34		MY	0	4
1074				48		MY	2	5
1080				57		Y	1	5
1081				46		MY	0	5
1102				19		MY	0	4
1107				32		Y	1	4
1108				28		MY	0	4
1110				26		N	0	4

Note. POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; OPT = oral production tests; EIT = elicited imitation tests; GJT = grammaticality judgment tests; CELT = the results of the structure section of the Comprehensive English Language Test for Learners of English (CELT Form A, Harris & Palmer, 1986); REC = the number of recasts provided in the treatment task; PREV = previous knowledge; MLT = the scores of the meta-language tests; EXIT = participant's awareness about the purpose of the study. Retained (that is, retained effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the posttest and maintained the improvement on the delayed posttest. Delayed (that is, delayed effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the delayed posttest. Short (that is, short-term effects) in the columns of OPT, EIT, and GJT refers to the participant who performed better on the posttest. Y = yes; N = no; MY = maybe yes.

APPENDIX U

KENDALL'S TAU-B CORRELATION COEFFICIENTS

OF THE CELT SCORES WITH THE DIFFERENCE SCORES

BETWEEN TWO OF THE THREE TESTS (ADJECTIVE ORDERING)

	POS (n = 18)		REC (n = 20)		REC+ (n = 16)		SAL (n = 19)	
	r_T	p	r_T	p	r_T	p	r_T	p
OPT								
Pre-Post	-.28	.145	-.08	.640	.34	.097	.02	.912
Pre-Delayed	-.06	.771	.01	.968	.16	.437	-.03	.883
Post-Delayed	.20	.296	.09	.623	-.08	.705	-.01	.943
EIT								
Pre-Post	-.05	.783	-.50	.004	-.27	.169	-.14	.428
Pre-Delayed	-.13	.471	-.24	.155	-.43	.027	-.24	.172
Post-Delayed	-.10	.587	.13	.441	-.24	.229	-.21	.242
GJT								
Pre-Post	.14	.444	-.37	.028	.11	.553	.01	.971
Pre-Delayed	.30	.096	-.21	.212	.08	.681	.19	.283
Post-Delayed	.11	.555	.23	.170	.08	.682	.05	.790

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; The shadowed cells indicate that probabilities are less than .05. CELT scores were not obtained from two participants of the POS group.

APPENDIX V

KENDALL'S TAU-B CORRELATION COEFFICIENTS

OF THE CELT SCORES WITH THE DIFFERENCE SCORES

BETWEEN TWO OF THE THREE TESTS (INDIRECT PASSIVES)

	POS (n = 19)		REC (n = 22)		REC+ (n = 27)		SAL (n = 20)	
	r_T	p	r_T	p	r_T	p	r_T	p
OPT								
Pre-Post	-.09	.618	-.22	.202	-.04	.777	-.18	.313
Pre-Delayed	.04	.838	-.11	.535	.04	.777	-.06	.762
Post-Delayed	.26	.168	.15	.389	.02	.913	.19	.311
EIT								
Pre-Post	-.41	.023	.01	.976	-.24	.099	-.40	.028
Pre-Delayed	-.28	.121	.07	.677	-.27	.069	-.23	.204
Post-Delayed	.21	.244	.08	.646	.06	.720	.17	.361
GJT								
Pre-Post	-.18	.276	.07	.669	.02	.867	-.15	.360
Pre-Delayed	-.10	.549	-.11	.477	-.14	.332	.01	.974
Post-Delayed	-.02	.911	-.16	.304	-.22	.119	.19	.264

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; The shadowed cells indicate that probabilities are less than .05. CELT scores were not obtained from two participants of the POS group. CELT scores were not obtained from two participants of the POS group.

APPENDIX W

KENDALL'S TAU-B CORRELATION COEFFICIENTS

OF THE NUMBERS OF RECASTS WITH THE DIFFERENCE SCORES

BETWEEN TWO OF THE THREE TESTS

	Adjective Ordering				Indirect Passives			
	REC (*n* = 20)		REC+ (*n* = 16)		REC (*n* = 22)		REC+ (*n* = 27)	
	r_T	*p*	r_T	*p*	r_T	*p*	r_T	*p*
OPT								
Pre-Post	-.14	.482	-.09	.679	.47	.011	.57	.000
Pre-Delayed	-.30	.148	.11	.606	.30	.120	.25	.124
Post-Delayed	-.16	.402	.05	.802	-.20	.299	-.20	.214
EIT								
Pre-Post	.34	.076	.50	.017	-.13	.479	.21	.190
Pre-Delayed	.07	.719	.19	.379	.05	.774	.19	.225
Post-Delayed	-.20	.291	-.17	.432	.14	.449	-.14	.403
GJT								
Pre-Post	-.04	.830	.35	.090	-.23	.187	.10	.514
Pre-Delayed	.01	.943	-.20	.332	.07	.690	.09	.570
Post-Delayed	.07	.696	-.65	.002	.27	.124	-.18	.230

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; REC = REC Group; REC+ = REC+ Group; The shadowed cells indicate that probabilities are less than .05.

APPENDIX X

KENDALL'S TAU-B CORRELATION COEFFICIENTS

OF THE PREVIOUS KNOWLEDGE WITH THE DIFFERENCE SCORES

BETWEEN TWO OF THE THREE TESTS (ADJECITVE ORDERING)

	POS (n = 20)		REC (n = 20)		REC+ (n = 16)		SAL (n = 19)	
	r_T	p	r_T	p	r_T	p	r_T	p
OPT								
Pre-Post	-.08	.680	.35	.089	.06	.794	.00	1.000
Pre-Delayed	.19	.363	.12	.566	.03	.876	-.13	.534
Post-Delayed	.26	.204	-.25	.213	-.13	.543	-.09	.656
EIT								
Pre-Post	-.24	.229	.05	.813	.13	.556	-.03	.871
Pre-Delayed	.06	.768	.11	.556	-.14	.522	-.25	.208
Post-Delayed	.21	.283	.03	.874	-.06	.766	-.39	.058
GJT								
Pre-Post	-.05	.805	.19	.330	.40	.057	.07	.716
Pre-Delayed	.13	.500	-.05	.785	-.03	.883	-.01	.968
Post-Delayed	-.01	.943	-.17	.352	-.45	.032	-.09	.667

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; The shadowed cells indicate that probabilities are less than .05.

APPENDIX Y

KENDALL'S TAU-B CORRELATION COEFFICIENTS

OF THE PREVIOUS KNOWLEDGE WITH THE DIFFERENCE SCORES

BETWEEN TWO OF THE THREE TESTS (INDIRECT PASSIVES)

	POS (*n* = 21)		REC (*n* = 22)		REC+ (*n* = 27)		SAL (*n* = 20)	
	r_τ	*p*	r_τ	*p*	r_τ	*p*	r_τ	*p*
OPT								
Pre-Post	-.49	.016	-.24	.218	-.36	.037	-.35	.090
Pre-Delayed	-.41	.046	-.17	.390	-.21	.227	-.35	.105
Post-Delayed	.30	.148	.12	.559	.09	.607	.04	.845
EIT								
Pre-Post	-.21	.287	.26	.180	-.36	.032	.14	.491
Pre-Delayed	-.05	.797	.03	.892	-.39	.024	.02	.936
Post-Delayed	.27	.172	-.19	.330	.06	.753	-.33	.128
GJT								
Pre-Post	-.21	.258	-.10	.577	-.19	.235	-.20	.305
Pre-Delayed	-.20	.272	-.18	.327	-.22	.172	-.21	.256
Post-Delayed	-.09	.630	.01	.948	.09	.602	-.04	.819

Note. Pre-Post = pretest-posttest period; Pre-Delayed = pretest-delayed posttest period; Post-Delayed = posttest-delayed posttest period; POS = POS Group; REC = REC Group; REC+ = REC+ Group; SAL = SAL Group; The shadowed cells indicate that probabilities are less than .05.