STRATEGIC PLANNING, RECASTS, NOTICING, AND L2 DEVELOPMENT

A dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By

Mika Hama, MME

Washington, DC
December 2, 2012
STRATEGIC PLANNING, RECASTS, NOTICING, AND L2 DEVELOPMENTS

Mika Hama, MME

Thesis Advisor: Alison Mackey, PhD.

Abstract

Since the mid-1990s, the link between recasts and L2 development has been extensively tested, and the results from those studies have largely demonstrated that recasts have a positive effect on L2 learning. With this firm support from previous empirical evidence, studies have begun to focus on how recasts assist learning and under what conditions they are more likely to bring about SLA. Among several factors, recent studies indicated that task variables may affect the extent to which learners might benefit from recasts. In order to extend this line of research, this dissertation examined whether a specific task variable, strategic planning (pre-task planning), helps with L2 development and whether it has an effect on noticing of target forms.

The participants were 50 university students learning Japanese and randomly divided into two groups: Planning Group and No Planning Group. Each of those groups was further divided into two smaller groups: either a group that completed three immediate posttests or one that participated in an immediate stimulated recall interview. All participants received recasts on Japanese pragmatic rules associated with Japanese give verbs during treatment task interactions. L2 development was measured by immediate posttests. Participants’ awareness was gauged using three introspective measures: think-aloud protocols, stimulated recall interviews, and an exit questionnaire.
The results showed that the No Planning Group was more accurate than the Planning Group across all immediate posttests, and the difference was statistically significant for one of the posttests. Regarding awareness, however, no difference was found between the two groups. The results are discussed using 1) participants’ comments from the think-alouds, stimulated recalls, and exit questionnaires and 2) participants’ perceptions of task difficulty ratings. This study also addressed learners’ thought processes during task performance to shed light on the results.
Acknowledgments

This dissertation would not have been possible without the support of so many people. Here, I would like to express my sincere thanks to every person who was part of my Ph.D. education.

First and foremost, I would like to thank my committee, Dr. Mackey, Dr. Leow, and Dr. Révész. Each member of my committee is a distinguished scholar with whom every Ph.D. student would dream to work. I was fortunate enough to study under them and to have them on my dissertation committee. More importantly, I was given opportunities to work on a research project with each of these distinguished professors during the course of my Ph.D. education. What I learned from them was enormous, and my education at Georgetown University was nothing but incredible because of these three individuals.

I also want to thank Japanese professors, Dr. Mori, Omori-sensee, Sato-sensee, and Dr. Tsujioka, whom I met in the East Asian Languages and Cultures Department at Georgetown University. Without their help, it would not have been possible to successfully collect data for my dissertation. But their help also went beyond my dissertation. They have given me opportunities to teach in Georgetown University’s Japanese program over the last several years. The teaching experiences were amazing for a student like me, who specialized in second language acquisition. They were simply wonderful co-workers with whom I loved working—every minute of it. They not only helped me to become an effective instructor but also provided me with a warm and
supportive environment where I could take refuge from my sometimes lonely and stressful dissertation process.

I would also like to thank all of my wonderful amazing friends. First, Dr. Rebecca Sachs—she was my unofficial mentor. From the very beginning of this dissertation process, Rebecca was there for me to offer advice. She is an amazing researcher, and I was very fortunate to have her insights on my dissertation. Sakiko Kajino—she was my intellectual problem solver. Whenever I was stuck in my dissertation, I went to her for help. She listened to me, and every time she helped to solve my problem. Dr. Natalia Jacobsen, Josephine Hsiao, and Patrick Collins—they were my unofficial therapists. Their continuous encouragement throughout my dissertation writing process was amazing. Without them, the process could have been, to say the least, unbearable or not even possible. Luke Amoroso—he was my patient statistician. Every time I seeked his assistance, he immediately responded to my requests and patiently listened to all my statistical confusions/problems. Jennifer Mixon—she is my American “sister” and my English tutor. She has known me since I arrived in the US 17 years ago. She is the only witness (aside from myself) to my beautiful 17-year journey in the US. She also has been patiently correcting my English whenever I asked her for editing work. I know she has corrected my incorrect English article usage at least millions times! I also want to thank other wonderful friends who made my life at Georgetown University a very, very happy one: Dr. Maymona Al Khalil, Dr. Geri Atanassova, Karen Bozicovich, Dr. Luis Cerezo, Dr. Rusan Chen, Trinh Dang, Justin Kelly, Dr. Jackie Lou, Motoko Takada.
They are all incredible human beings. They are my treasures, and I am tremendously lucky to have them in my life.

Finally, special thanks go to my beloved family. There are no words to express my appreciation for what they have done for me during my Ph.D. process. I dedicate my dissertation and every accomplish I made to my family. I feel very blessed because I have such a wonderful family who unconditionally loves me and believes in me. I love them dearly.
This dissertation is dedicated to my family: My mom, dad, and sisters.
Table of Contents

Chapter 1: Review of Literature ........................................................................ 1
  Attention, detection, and noticing ................................................................. 1
  Theoretical arguments .................................................................................... 1
    Russell Tomlin and Victor Villa ................................................................. 1
    Richard Schmidt ....................................................................................... 3
    Peter Robinson ......................................................................................... 7
  Measuring noticing ...................................................................................... 8
  Think-alouds ............................................................................................... 10
  Stimulated recalls ...................................................................................... 11
  The importance of triangulation .................................................................. 14
  Empirical studies ....................................................................................... 15
    Studies demonstrating unaware learning .............................................. 15
    Studies not finding evidence of learning without awareness .............. 19
    Comparison of previous studies ........................................................... 23
  Summary of Noticing ................................................................................. 24
  Recasts ........................................................................................................ 25
    Theoretical arguments ............................................................................. 25
    Recasts and L2 development: Empirical Evidence ................................ 28
      Types of recasts .................................................................................... 29
      Target of recasts .................................................................................. 30
      Developmental readiness and proficiency ........................................... 32
      Individual differences .......................................................................... 33
      Task variables ...................................................................................... 36
  Recasts and noticing: Empirical evidence ................................................. 38
    Types of recasts .................................................................................... 39
    Target of recasts .................................................................................... 40
    Proficiency .............................................................................................. 40
    Individual differences ............................................................................ 41
Summary on recasts ................................................................. 42
Task planning and task performance ........................................ 42
Types of task planning ............................................................. 42
Measures of task performance .................................................... 43
Strategic planning ...................................................................... 44
   Two theoretical models .......................................................... 44
   Levelt’s model of speech production ......................................... 46
Empirical studies ...................................................................... 47
   Amount of planning time ......................................................... 48
   Linguistic features .................................................................. 49
   Proficiency .............................................................................. 50
   Type of tasks used ................................................................... 52
   Length of task performance ..................................................... 53
   Availability of within-task planning .......................................... 54
   Planning instructions .............................................................. 55
   Learners’ perceptions ............................................................. 57
Summary of task planning and task performance ....................... 59
Strategic planning, noticing, and L2 development ....................... 59
   Issues with previous strategic planning studies on task performance .... 59
   Implications from the two theoretical models of attention ............. 61
   Empirical study ...................................................................... 64
   Summary on strategic planning, noticing, and L2 development ....... 67
Research questions ................................................................... 67

Chapter 2: Method ..................................................................... 70
   Participant .............................................................................. 70
   Linguistic target ..................................................................... 71
      Scope of the linguistic target addressed in the experiment .......... 75
   Participant criteria ................................................................. 76
Chapter 3: Results .................................................. 114
  Research question 1 .................................................. 114
    Overall results .................................................. 115
      Pretests ......................................................... 115
      Oral tests ..................................................... 116
      Written production tests ..................................... 117
      Written recognition tests ................................. 119
    Summary of three test results ............................. 121
  Ageru and kureru posttest results ....................... 122
    Oral posttest ................................................ 122
    Written production posttest ............................... 123
    Written recognition posttest ............................... 126
  Ageru and kureru frequency of usage during posttests .. 127
    Summary of ageru and kureru posttest results .......... 130
  Results from the three test-item types ................... 130
    Oral posttest ................................................ 130
    Written production posttest ............................... 133
    Written recognition posttest ............................... 135
    Summary of results from the three test-item types .. 137
  Research question 2 ............................................ 138

Chapter 4: Discussion ............................................ 140
  Research question 1 ............................................ 142
    Prediction from the two models of attention ............ 142
    Results from this study ..................................... 142
Discussion of the results ......................................................... 143
  Comments from stimulated recalls and exit questionnaires .......... 143
  Participants’ perception of task difficulties ............................. 150
  Content of planning and L2 development ............................... 153
Research question 2 ............................................................... 157
  Results from this study ...................................................... 157
  Possible limitations of awareness measurements ..................... 158
  Contextual factors and pragmatic principle ........................... 161
Limitations ............................................................... 165
  Tasks and procedures ..................................................... 165
  Overall design of the study .............................................. 170
Conclusion ................................................................ 171

Appendices ................................................................... 174
  Appendix A: Consent form ................................................. 174
  Appendix B: Instructions and trial items for oral tasks .............. 178
  Appendix C: List of people who appeared on the oral tasks ...... 187
  Appendix D: Instructions for written tests ............................. 188
  Appendix E: Think-alouds practice task ............................... 189
  Appendix F: Treatment task instructions for the No-Planning Groups .............. 190
  Appendix G: Stimulated recall instructions ........................... 191
  Appendix H: Biodata questionnaire ....................................... 192
  Appendix I: Treatment tasks ............................................. 193
  Appendix J: Strategic planning instructions for the Planning Groups .............. 207
  Appendix K: Oral pretest .................................................. 208
  Appendix L: Written production pretest items ......................... 213
  Appendix M: Written recognition pretest items ....................... 216
  Appendix N: Oral posttest ................................................ 219
Appendix O: Written production posttest items ............................... 226
Appendix P: Written recognition posttest items ............................... 230
Appendix Q: Questionnaire .......................................................... 234
Appendix R: Noticing from four data sources ..................................... 236

References .................................................................................. 237
List of Figures

*Figure 1.* Implications from the two models of attention ........................................ 64

*Figure 2.* *Ageru* and *kureru* in relation to in-group/out-group .......................... 73

*Figure 3.* Experimental design .................................................................................. 77

*Figure 4.* Mean percentage of correct answers in oral pretest and posttest
by group ...................................................................................................................... 117

*Figure 5.* Mean percentage of correct answers in written production pretest and
posttest by group ......................................................................................................... 119

*Figure 6.* Mean percentage of correct answers in written recognition pretest and
posttest by group ......................................................................................................... 121

*Figure 7.* Mean oral posttest score for each verb by group ..................................... 123

*Figure 8.* Mean written production posttest scores for each verb by group .......... 125

*Figure 9.* Mean written recognition posttest scores for each verb by group .......... 127
List of Tables

Table 1. Time pilot participants used for strategic planning ......................... 94
Table 2. Coding scheme of noticing .......................................................... 109
Table 3. Coding from the two coders on noticing of form ......................... 111
Table 4. Coding from the two coders on noticing of contextual factors ....... 112
Table 5. Pretests accuracy by group ......................................................... 115
Table 6. Oral pretest and posttest accuracy by group ............................... 116
Table 7. Written production pretest and posttest accuracy by group ......... 118
Table 8. Written recognition pretest and posttest accuracy by group .......... 120
Table 9. Oral posttest scores for each verb by group ................................. 122
Table 10. Written production posttest scores for each verb by group ........ 124
Table 11. Written recognition posttest scores for each verb by group .......... 126
Table 12. Numbers of participants who used only ageru during each posttest .... 128
Table 13. Frequencies of verb use in each posttest by group ...................... 129
Table 14. Oral posttest scores of test item types by group ......................... 131
Table 15. Oral posttest scores for each verb by test-item types and group .... 132
Table 16. Written production posttest scores for test-item types by group .... 133
Table 17. Written production posttest scores for test-item types by verb and group ................................................................. 134
Table 18. Written recognition posttest scores of test-item types by group .... 135
Table 19. Written recognition posttest scores of test-item types by verb and group ................................................................. 137
Table 20. Number of participants who noticed contextual factors/pragmatic principle by types and group .................................................. 139
Table 21. Posttest accuracy and participant’s reports.................................. 149
Table 22. Ratings of perceived task difficulty .......................................... 150
Table 23. Test improvement and content of planning ............................... 156
Chapter 1: Review of Literature

*Attention, Detection, and Noticing*

SLA researchers who follow cognitive linguistic paradigms such as the interaction approach consider attention and noticing to be important cognitive processes for L2 development. These concepts are claimed to be crucial for understanding how L2 input is processed and becomes available for further processing necessary for learning. In the following, I will review publications from the key researchers who have posed theoretical claims about attention and noticing, and their roles in L2 development: Tomlin and Villa, Schmidt, and Robinson.

*Theoretical Arguments*

*Russell Tomlin and Victor Villa.* Tomlin and Villa (1994) argued that attention is necessary for SLA. Based on the work of Posner (1992) in the field of cognitive science, they argued that attention is comprised of a variety of mechanisms, not a unitary phenomenon; therefore, a finer-grained look at attention is needed in order to understand how attention is committed to a task or input, and its role in SLA. They stated that attention has at least three separate principal components: alertness, orientation, and detection. Those functions are claimed to be interrelated but can also operate independently.

Regarding the attentional function of alertness, Tomlin and Villa stated that it “represents an overall, general readiness to deal with incoming stimuli or data” (p. 190). More simply, alertness may be associated with, but is not limited to, aspects such as motivation and interest in the L2. A learner’s alertness level is said to influence his/her
processing or performing of a task. When there is an increase in alertness, there is also a possible increase in the rate at which information is selected for further processing. Alertness can either work separately from or modulate the second function of attention, orientation.

Orientation, or directing attentional resources toward some type of sensory stimuli, Tomlin and Villa assert, can have either facilitative or inhibitory consequences for further processing. Stimuli toward which attention is oriented are likely to be cognitively registered, or have activation of another attentional function, detection. On the other hand, stimuli which do not receive the specific aligning of attention are said to be inhibited; thus, their detection becomes more effortful.

The third attentional function, and the one argued by Tomlin and Villa to be crucial for learning, is detection, which refers to “the cognitive registration of some stimuli” (p. 190). Tomlin and Villa stated that detection is “the process by which particular exemplars are registered in memory and therefore could be made accessible to whatever the key processes are for learning, such as hypothesis formation and testing” (p. 193). This detection can be aided by the previously mentioned attentional functions, alertness and orientation. When a learner is more alert and/or oriented to particular stimuli, detection is likely to occur. However, neither alertness nor orientation is required for the occurrence of detection. Crucially, they argued, detection does not need to be accompanied by awareness, though they acknowledged that awareness does have a facilitative effect in creating situations allowing detection to occur.
Richard Schmidt. Schmidt (e.g., 1990, 1993a, 1993b, 1994, 1995, 2001, 2010) is indisputably the most well-cited scholar regarding the topic of attention and awareness, and their roles in SLA. The important difference between Schmidt’s and Tomlin and Villa’s arguments is that noticing, which entails awareness, is an essential factor in SLA.

In his original Noticing Hypothesis (e.g., 1990, 1993a, 1993b), he argued that noticing is an “allocation of attentional resources to some stimulus and identifies the level at which perceived events are subjectively experienced and are reportable by the person which experiences them” (Schmidt, 1993b, p.24), and noticing is a necessary and sufficient condition for input to become intake (i.e., the part of input internalized by a learner). Noticing is the result of an attentional mechanism that Schmidt (1995) has proposed to be “nearly isomorphic with attention” (p. 1). Schmidt further stated that noticing entails conscious registration of “elements of the surface structure of utterances in the input-instances of language, rather than any abstract rules or principles of which such instances may be exemplars” (2001, p. 5). In other words, noticing represents a lower level of awareness, whereas awareness of a rule or generalization, which Schmidt termed understanding, represents a higher level of awareness. Understanding, he argues, is not necessary for intake but is facilitative of L2 development.

Schmidt’s original Noticing Hypothesis of no learning without awareness was a very strong claim, which he later acknowledged to be both non-falsifiable (see, Truscott, 1998) and methodologically impossible to verify, as noticing is fleeting and cannot be completely recorded. The zero-point question, Schmidt (1994, 1995, 2001) further states, is theoretically interesting but has little practical value. For these reasons, he later
modified the Noticing Hypothesis (1994, 1995, 2001, 2010), to make the weaker claim that noticing has facilitative effects on SLA and that more noticing is associated with greater learning. Even while acknowledging that his strong claim was not falsifiable, Schmidt still concluded that noticing is necessary to attain robust memory; a review of previous research in cognitive psychology suggested universal agreement on that point. Kellogg and Dare (1989), for instance, argued that “the degree of elaboration resulting from unattended encoding appears to be too limited to have any substantive influence on human cognition or behavior” (p. 412). Schmidt (2001) also continued to claim that noticing may be necessary for successful L2 learning in certain situations, considering that “many features of L2 input are likely to be infrequent, non-salient, and communicatively redundant, intentionally focused attention may be a practical (though not theoretical) necessity for successful language learning” (p. 21).

Schmidt further elaborated on what learners need to notice in order to develop various aspects of an L2. He stated, “what must be attended to and noticed is not just the input in a global sense but whatever features of the input are relevant for the target system” (1993a, p. 209). This means that learners must attend to the sounds of language input in order to acquire phonology. For syntax and semantic acquisition, they must attend to the word order and the meaning of the word. In terms of pragmatics, they must attend to both linguistic form and the relevant associated social and contextual features (1990, 1993b, 1995). Additionally, Schmidt theorized that attentional requirements, including the degree of focal attention required, may differ according to the area of language involved (e.g., morphology, syntax). Whereas it may be relatively
straightforward for learners to learn the morphological expression of past tense in an L2 if their L1 also expresses past tense morphologically, other areas of language, such as pragmatics, may be less transparent. For example, Schmidt (1993b) stated, “learning a new pragmatic system often entails learning how to make a new interpretative assessment of the world” (p. 34). In other words, “the relevant contextual factors to be noticed are likely to be defined differently or may be nonsalient for the learners” (p. 36). Therefore, “simple exposure to sociolinguistically appropriate input is unlikely to be sufficient” (p. 36) for noticing to occur and deliberate focal attention may be necessary.

Whether learners notice input or not, Schmidt stated, is determined by many factors; therefore, learners “are not free to notice whatever they want” (1990, p. 144). The factors argued to influence noticing include, but are not limited to, the following: 1) interpretability; 2) L1, L2, and prior knowledge; 3) frequency and perceptual saliency; 4) meaningfulness of the input; 5) proficiency level; 6) learner-internal factors such as working memory and motivation; and 7) task demands. More specifically, input is likely to be noticed when it is interpretable (e.g., VanPatten, 1996); meets learners’ expectations or relates to their L1, L2, or prior knowledge (e.g., Carroll, 1999; Ellis, 1994; Gass, 1997; Park, 2010); is frequent or perceptually salient (e.g., Bardovi-Harlig, 1987; Ellis, 1994; Gass, 1997; Mackey, 2006); or is meaningful or necessary for understanding (e.g., Mackey, Gass, & McDonough, 2000; VanPatten, 1996; Mackey, 2006). As far as developmental level is concerned, beginning learners may not notice as much as advanced learners (Atannassova, 2012); and in addition to the higher developmental level, individual differences, such as higher working memory capacities and levels of
motivation, are also argued to have positive effects on the noticeability of input (e.g., Al-Khalil, 2011; Mackey, Philp, Fujii, Egi, & Tatsumi, 2002; Robinson, 2001a; Robinson, 2003).

Tasks learners engage in classrooms and elsewhere are also an important factor to consider. What learners notice may be manipulated by the demands of the tasks in which they engage (Schmidt, 1990; Robinson, 2010; Robinson & Gilabert, 2007). Schmidt (1990) also proposed that task instructions may possibly direct learners’ attention to certain aspects of the input they receive and increase noticing of certain forms in subsequent input. For example, when an instruction directs learners to attend to certain features of the target language, those features become more salient, and therefore, facilitates noticing of those features. Schmidt supports the idea that task instructions and classroom instructions can have positive effects on noticing, but at the same time, he cautions against regarding them as a guaranteed means of directing learners’ attention and thus promoting the noticing of specific elements in input. He argues that we cannot eliminate the possibilities that learners may fail to understand grammar explanations provided in class, or that they will focus, either deliberately or non-deliberately, on items which the task instructions did not require. Certainly also, an intention to attend to certain aspects of the input does not assure noticing of those aspects, due to various constraints mentioned previously which are outside of learners’ control (e.g., developmental level, working memory). However, there is some degree of autonomy in the focus of learners’ attention, and Schmidt acknowledges that deliberate attention may have a facilitative effect on noticing.
Peter Robinson. In a review paper, Robinson (1995) reconciled the two different positions argued by Tomlin and Villa (1994) and Schmidt (1990, 1993) by presenting a model of the relationship between attention and memory. By defining noticing as “detection plus rehearsal in short-term memory” (1995, p. 296), Robinson placed detection, which occurs without awareness, as a precursor to noticing, which includes awareness. He considered awareness to be a consequence of rehearsal and, like Schmidt, claimed it was necessary for learning.

Robinson (2003) explained,

[S]hort-term memory is a part of long-term memory in a currently heightened state of activation. Long-term memory is where instances of encoded input are stored and assume… the representational shape that recognition processes match to new instances of input in working memory during parsing and comprehension (p. 631).

Working memory, in turn, is “a subset of short-term memory” (p. 631), or focal attention which coincides with awareness. Robinson argued that sensory information which is detected and pre-attentively processed may briefly enter one’s passive short-term memory (as opposed to active short-term memory or working memory) and access information previously encoded in long-term memory outside of awareness. However, “For newly detected information to be encoded in long-term memory…to be learned, the information must enter focal attention and so short-term working memory, where rehearsal processes operate prior to encoding in long-term memory” (2003, p. 654) or L2 learning. In addition, “permanent encoding in long-term memory is a consequence of the
level of activation of information in short-term memory, itself the result of rehearsal and elaboration” (1995, p. 298). Simply put, Robinson claimed that in order for new information that has never been encoded in long-term memory to be learned, detection itself is insufficient, and some level of awareness is necessary—concurring with Schmidt’s stance. Robinson also acknowledged, similar to Schmidt, the effects of learner-internal factors on noticing. Robinson (1995, 2003) noted that the amount learners can notice and subsequently rehearse in short-term memory while performing a task may be affected by their attentional and memory resources.

To summarize, the theoretical claims described above do not reach consensus regarding whether or not awareness is necessary for SLA. One thing, however, that appears to be mutually agreed amongst them is that noticing, or awareness, has facilitative effects on SLA. Schmidt and Robinson argued that learners’ noticing may be affected by a number of factors. For example, in terms of the area of language involved, the degree of focal attention required for noticing to occur may differ. In other words, while simple exposure to input suffices for the noticing of the morphological expression of past tense, deliberate focal attention may be necessary for the noticing of new pragmatic systems. Task demand and other factors associated with tasks is also claimed to change the way learners notice. These points will be further addressed in this dissertation.

Measuring Noticing

In order to examine whether or not awareness is necessary for SLA and/or in what conditions noticing is likely to occur, researchers have been measuring learners’ noticing
by employing introspective techniques. Those techniques can be classified as 1) online
measures, which involve collecting data while learners are interacting with L2 input, and
2) offline measures, which involve collecting data after learners are exposed to L2 input.
Online sources of data include eye tracking (Godfroid, Housen, & Boer, 2010), notes
taken by learners (e.g., Hanaoka, 2007; Izumi, 2002), recordings of private speech (e.g.,
Ohta, 2000), and think-aloud protocols (e.g., Alanen, 1995; Hama & Leow, 2010;
Jourdenais et al., 1995; Leow, 1997, 2000, 2001; Rosa & O’Neill, 1999; Sachs & Suh,
2007). Offline sources include diaries (e.g., Schmidt & Frota, 1986), questionnaires (e.g.,
Robinson, 1997a, b), interviews (e.g., Williams, 2005), immediate recall (Philp, 2003),
immediate reports (Egi, 2004), and stimulated recall protocols (e.g., Adams, 2003; Egi,
2010; Mackey, 2006; Mackey et al., 2000; Nabei & Swain, 2002; Swain & Lapkin, 2002).
Among those various measures, two types of verbal protocols, think-aloud and stimulated
recall, have been the most widely utilized in SLA.

Verbal protocols, if conducted appropriately, can be reliable tools to gain insight
into learners’ cognitive processes. Ericsson and Simon (1984, 1993) have argued that
information focally attended in short-term memory (i.e., in working memory) is
retrievable through verbal reports, and under certain conditions verbalizations can be
regarded as direct representations of learners’ thought processes. Not only did they
categorize verbal reports based on time frames--concurrent (i.e., online) and retrospective
(i.e., offline), as described above--but they also differentiated among the types of
reporting, such as simple verbalizations of thoughts, or verbalization of explanations or
justifications. Depending on these factors, they argued, the validity of reports would change.

Think-alouds. In think-aloud protocols, or concurrent verbal reports, learners are instructed to verbalize the thoughts that come to mind as they complete a task. Thus, the protocols are designed to tap into learners’ short-term memory and to access the information which they have just attended. The absence of a time lag between the exposure to L2 data and the reporting can reduce the possibility of memory decay; therefore, it increases the validity of the data. This concurrency of reporting is an obvious advantage; however, it has also been argued to create a risk of reactivity, which may threaten the validity of the verbalization data (e.g., Ellis, 2001; Jourdenais, 2001, Robinson, Mackey, Gass, & Schmidt, 2012). For example, engaging in concurrent verbal reporting while completing a task may function as an additional task, thereby, imposing an extra processing load, and may change the task performance itself. As a result, the learners’ cognitive processing, which researchers are attempting to capture, may be altered.

Ericsson and Simon (1984, 1993) argued that whether or not think-alouds are reactive depends on the type of verbalization made by participants. Their claim was largely supported by findings from studies in cognitive psychology (see also Bowles, 2010, for a review of those studies). The studies indicated that when participants verbalized explanations or justifications of their thoughts, think-alouds tended to be reactive, affecting the participants’ cognitive processes and task performance. However,
when they were simple verbalizations of participants’ thoughts per se, they were almost consistently non-reactive.

Although the findings from the studies in cognitive psychology were informative regarding the reactivity issue, Bowles (2010) argued, many used non-verbal tasks, and their relevance for SLA cannot be assumed. Bowles, therefore, carried out a meta-analysis to investigate the issue of reactivity of think-alouds in verbal tasks by analyzing relevant studies from both cognitive psychology and SLA. Based on the general effect sizes for those studies, she stated the use of think-alouds, while completing a verbal task, has only a small effect on task performance, whether in a positive or a negative direction. In addition, Bowles found that “in 86 percent of the effect size calculations, the 95 percent confidence interval overlaps zero” (p. 138); that is, most of the time, it wasn’t clear that there was a reliable effect. Bowles interpreted this as suggesting that “verbal reports can reliably be used as a data collection tool” (p. 138). In a subsequent analysis, she found several factors which accounted for the between-study variances, including 1) type of verbalization, which has been extensively studied in cognitive psychology, 2) learners’ L2 proficiency, and 3) the explicitness of the directions in treatment tasks. For instance, thinking aloud during grammar learning tasks had different effects on learners’ L2 performance depending on whether the treatments were implicit or explicit. Bowles urged the systematic examination of these factors in order to ascertain their influence on reactivity.

*Stimulated recalls.* Along with think-aloud protocols, stimulated recall protocols have also been used quite extensively in SLA. Think-aloud protocols, concurrent verbal
reports, are more ideal measures of noticing than stimulated recall, retrospective verbal reports, because think-alouds carry a limited threat to the validity of the data resulting from the time lag between the exposures to L2 data and the reporting. However, thing-aloud protocols are not feasible for all types of data collection, especially for research that focuses on observing learners’ cognitive processes during oral communicative activities because of the impossibility of engaging in conversation while simultaneously verbalizing one’s inner thoughts (e.g., Bowles, 2010; Jourdenais, 2001) and stimulated recall protocols have been a preferred method of data collection. In stimulated recall protocols, learners first engage in L2 activity. These learners’ task performances are video- and/or audio-recorded for a later recall session. Upon the completion of the activity, the learners and a researcher view and/or listen to (parts of) the recording. With the help of these visual and auditory stimuli or cues, learners are instructed to verbally recall their thoughts at the time of the activity.

As mentioned earlier, a major concern regarding retrospective measures is memory deterioration. Because there is a time lag between the point of cognitive processing and that of reporting, it has been warned that this type of verbalization may not be veridical; that is, it may not accurately represent the actual cognitive processing. Due to the time lag, learners may forget their original thoughts; therefore, their reports may be incomplete. Alternatively, they may report what they think they were thinking at the time of the activity—a reconstruction of their thoughts which may not be in accord with their original thoughts. In the case of stimulated recall protocols, however, veridicality concerns resulting from memory decay are reduced with the use of video- or
audio-recordings as stimuli to serve as memory aids and assist with recall. This allows researchers to obtain more accurate and complete responses (Bloom, 1954; Gass & Mackey, 2000; Mackey, 2012; Mackey, Abbuhl, & Gass, 2012).

In regard to reactivity, it has been stated that it is less of a concern with stimulated recall protocols than with think-aloud protocols (e.g., Gass & Mackey, 2000). Indeed, it cannot affect performance on the task which participants are recalling because the task has already been completed before the recall. However, there have been reports that stimulated recall may be reactive on tasks performed following the recall session (e.g., Adams, 2003; Egi, 2004 however, see Egi, 2008). This may occur because during the recall session, learners are exposed to the same L2 input they experienced in the initial activity; therefore, it creates a situation called double input-exposure (Leow, 2002). This means that learners have an additional opportunity for processing input which they may have not been able to process during the initial exposure. Additionally, this double exposure may possibly present a threat to the veridicality of reports utilizing stimulated recall: Learners may select to report the thoughts they have during the second exposure to L2 data instead of the thoughts they had previously during the original exposure. With said concerns in mind, it has been suggested that stimulated recall protocols should be carefully implemented and their data must be interpreted with care (Gass & Mackey, 2000).

In order to increase the validity of reports from stimulated recall protocols, it has been suggested that researchers should consider several factors. These include, but are not limited to: timing, cue strength, instructions to the participants during recall, and what
to do or not to do in a stimulated recall interview (e.g., Egi, 2004, 2008; Gass & Mackey, 2000; Mackey, 2012; Jourdenais, 2001). First and foremost, researchers need to ensure that the time lag between the recall and the event to be recalled is kept as minimal as possible. If recall sessions are not conducted immediately after the completion of a task, it can result in significant memory decay, and learners may have to rely on what they think happened. Second, the cue strengths must be sufficient to elicit accurate and complete responses. Combinations of stimuli, such as visual and audio, are more ideal than a single source of stimuli. Third, the instructions given to learners during recall sessions must clearly state that the learners must report only what they were thinking at the time of the original task performance. Researchers should also constantly remind learners to focus on recalling the thoughts they had during the first exposure to the L2 input and not on the thoughts that may occur during the recall sessions. Fourth, if the recall session occurs in a one-on-one interview setting, researchers must refrain from asking for reasons or explanations during the interviews, as doing so can deter learners from focusing on recalling their original thoughts and may lead to fabrication of thoughts. In addition, in order to avoid an interviewer effect or an effect related to the socially situated nature of the interview, which may drive learners to report what they think their interviewer wants to hear instead of their actual thoughts, researchers must act as passive listeners and demonstrate that social interaction is not intended during the interviews.

The importance of triangulation. As with all measures utilized in research, think-aloud protocols and stimulated recall protocols cannot be considered flawless measures (Mackey, 2012; Robinson et al, 2012). Researchers have called for further examinations
of these methods and, in particular, of the conditions in which verbal reports are less likely to suffer from validity concerns (e.g., Bowles, 2010; Egi, 2008). Jourdenais queries “whether or not learners are actually able to verbalize everything that they are thinking” and states that “the comfort level of learners with this type of task (as with any other) and their ability to provide protocols must be considered” (p. 374). It has also been cautioned that the lack of verbal report cannot be taken as evidence that learners did not notice something (e.g., Allport, 1988; Mackey et al., 2000; Rosa & O’Neill, 1999). In many studies, therefore, in order to increase the robustness of their data, researchers acquire triangulated data by additionally utilizing some type of offline measures (see Hama & Leow, 2010; Leow, 1997; Mackey, 2006).

In conclusion, keeping all of the concerns related to think-alouds and stimulated recalls in mind, it still holds true that they are the measures which have been the most examined and accepted in SLA. When the previously mentioned methodological precautions are taken, researchers can significantly diminish the concerns regarding verbal reports and can collect valid and reliable noticing data.

*Empirical Studies*

There is a series of six empirical studies in which a set of researchers have debated theoretical arguments related to noticing and awareness. The results from these studies have been mixed; some showed awareness is not necessary for learning, and others contradicted those results.

*Studies demonstrating unaware learning.* Williams (2004) examined whether or not participants were able to learn a feature of miniature noun class systems, specifically
animacy rules, without awareness. Williams used a noun class system which included artificial nouns and eight artificial determiners whose surface forms resembled those of Italian. The determiners were governed by three features: definiteness, number, and animacy. Prior to the training task, the participants learned the meanings of artificial nouns and the definiteness features of the determiners. The number and animacy features were not disclosed. During the training task, the participants were auditorily presented with determiner-noun combinations, or noun phrases. For each of the noun phrases, the participants orally repeated it, indicated whether it referred to a living or non-living thing, and translated it to English. In the following test task, they were presented with English translations of the noun phrases on a computer. For each noun phrase in English, the participants were asked to select a noun phrase in the artificial language from two alternatives which were provided. Upon the completion of the test task, in an retrospective interview, the participants were asked what criteria they had utilized to make their selections on test items. Seven of the 37 participants reported the animacy rules, and the remaining participants were labeled as unaware learners. On generalization items (that is, items containing determiner-noun combinations which had not occurred during the training), unaware learners’ performance was 61% accurate, significantly above the chance level of 50%. This result provides support to theories of learning without awareness. Williams, however, also reported that participants who spoke an L1 in which gender is marked morphologically or knew many gender languages produced significantly more accurate answers compared to those who did not, and that the performance of participants who did not speak a gendered language was not significantly
different from chance. Williams concluded that because the determiners were created based on Italian, prior knowledge may have indirectly influenced the results.

Following his 2004 study, Williams (2005) conducted two additional experiments in an attempt to answer the same research questions as his 2004 study. In these new experiments, Williams used a noun class system which included only four artificial determiners, *gi, ro, ul, ne*, which carried two features: distance (near vs. far) and animacy. Prior to the training task, the participants learned the meanings of the determiners with regard to distance, but their animacy features were not disclosed. During the training, for each treatment item, the participants listened to a sentence which contained a target noun phrase, indicated whether it meant near or far, and repeated the sentence aloud. In the test task, the participants were presented with a sentence on a computer screen with a blank space where a noun phrase was to be supplied to complete the sentence. They were then asked to select one noun phrase from two alternatives which were given. The same awareness measure utilized in Williams’ 2004 study, an offline interview, was used in this study, and he found that eight of the 41 learners were able to correctly state the animacy rules. The results on generalization items showed that the unaware learners who did not report animacy rules nonetheless performed significantly more accurately than chance. The same results were found in his second experiment, in which each noun only occurred with one possible determiner during training. This modification was made to ensure the results were indeed due to the fact that participants learned animacy rules, not just that they developed some associations between determiners. With these results, Williams concluded that his study provided evidence for learning without awareness.
Leung and Williams’ 2011 study utilized the same surface forms of the four artificial determiners as Williams (2005), gi, ro, ul, ne; however, those determiners carried the features of thematic role (agent vs. patient) and age (adult vs. child). Prior to the training phase, participants learned one of the functions of the determiners, age, but not the function of thematic role. During the training phase, first, they were presented with a picture of two people on a computer screen and were asked to describe it in their own words. Second, they listened to a sentence with an artificial determiner that described the picture. The sentence structure did not always follow that of English, that is, half of the time, patient nouns appeared prior to agent nouns in sentences. As they listened to the sentence, they were asked to click right or left on the mouse as quickly as possible to indicate where on the screen the individual mentioned in the sentence was located. While it was possible to make a correct decision by listening to the entire sentence, it was designed in such a way that participants could decrease their response times by learning the thematic role features of the determiners. Third, the participants were asked to rearrange the sentence they heard into the correct order for English. In the test phase, the same procedure was followed, and the participants’ reaction times were used in order to determine whether they had learned the thematic features of the determiners. Some of the test items were violation items; that is, they violated the thematic rules to which the participants had been exposed. After the test phase, in a retrospective interview, the participants were shown some of the training items and were asked “if they had any feelings about when gi versus ro and ul versus ne were used” (p. 45). In the words of Leung and Williams, “participants were classified as being aware or
unaware on the basis of their reported knowledge of the target form-meaning mapping” (p. 45). Or, aware participants were the ones who “were able to link the use of articles [determiners] with concept such as someone doing something to the other person or one person takes as active role and the other is more passive” (p. 46). The result from the unaware learners showed that their reaction times decreased throughout the experiment, especially during the training phase. In addition, there were a few attempts among the unaware learners to respond in less than one second, which could be accomplished only by having learned the determiner’s thematic role. The most important finding, according to the researchers, was that participants’ reaction times were significantly slower on the test’s violation items, or, as the researchers put it, “their reaction times were affected when this [learned] critical association was interrupted” (p. 49). Considering this finding together with those of previous studies on the topic, Leung and Williams claimed that their study provided additional proof of learning without awareness.

*Studies not finding evidence of learning without awareness.* Leow (2000) addressed the relationship between awareness, or a lack thereof, and L2 learners’ subsequent recognition and production of Spanish irregular stem-changing verbs. Adults learning Spanish as a foreign language participated in a problem-solving task, specifically a crossword puzzle. In the puzzle the correct target forms were already provided; therefore, participants were given opportunities to notice the target forms. However, because this study’s focus was to elicit data from both aware and unaware learners, this puzzle was designed not to deliberately bring participants’ attention to the target forms. The puzzle directed participants’ attention to other forms, verb endings. In other words,
learners had to fill in the correct endings of the verbs for the preterit tense in Spanish to complete the puzzle. While completing the crossword puzzle, participants were instructed to think aloud. Immediately upon completion of the crossword puzzle, participants engaged in a written fill-in-the-blank production task in which they were asked to produce the target forms in different contexts. The items used in the task were identical to those in the crossword clues. The participants then completed a four-choice recognition task. During the assessment tasks, participants were also instructed to think aloud. Their verbal reports were then coded either “aware” or “unaware”. Participants were classified as aware if they “provided a report of being aware of the target forms [a simple reference to the target forms which does not require mentioning of rules] or some form of metalinguistic description of the underlying rule” (p. 564). All remaining participants were classified as unaware. Awareness was also cross-checked with offline awareness measures: a post-exposure questionnaire and an interview which occurred after the experiment. The scores of the aware participants’ recognition and written production tasks following the crossword task showed significant gains over their corresponding pretest scores. Conversely, among unaware learners, there was no significant difference between pretest and posttest performances. Leow concluded that the results supported the claim that awareness during initial processing plays an essential role in the subsequent recognition and production of L2 data.

Because of the contradicting results reported in Leow (2000) and Williams (2004, 2005), Hama and Leow (2010) revisited Williams (2005) to reexamine his results. Hama and Leow’s investigation was a conceptual replication of Williams’ study with several
changes. First, in addition to the offline post-exposure questionnaire, think-alouds were used during the training and testing phases as online measures of awareness. This change was made to increase the validity of the awareness data. Second, the number of options in the multiple choice assessment task was increased from two to four to decrease the influence of learners’ guesses. Third, another assessment task, a production task, was implemented as an additional assessment tool. Fourth, a finer-grained awareness coding scheme was adapted from Rosa and O’Neill (1999) and a verbal report was coded as: “(a) noticing when some aspect of animacy was mentioned or commented upon, (b) understanding when correct rules related to animacy were mentioned, or (c) no report when the report did not fall under the coding categories of noticing or understanding” (Hama & Leow, p. 13). Fifth, the modality of item presentation between the training and testing phases was matched, and both were completed orally. In Williams’s (2005), assessment tasks were provided in written mode, whereas the sentences in the training task were provided orally. Other than those changes, the procedures and materials generally followed that of Williams’ (2005) Experiment 2.

Verbal reports from the think-alouds revealed that at the level of noticing, there were five participants who provided either a partial report or an incorrect rule of assigning animacy to the novel words. At the level of understanding, there were three participants who made reference to the correct underlying animacy rules. In regard to the comments on the post-exposure questionnaire, three out of the five participants who had reported on some aspect of animacy at the level of noticing in their think-alouds also referenced either correct or incorrect animacy on the offline measure. The remaining two
participants did not mention animacy on their questionnaires. Additionally, one participant reported awareness of animacy at the level of noticing on the post-exposure questionnaire but not during the think-aloud protocol. At the level of understanding, all three participants who had verbalized correct animacy rules in their think-alouds also mentioned the correct rules on their questionnaires.

Excluding the learners who demonstrated awareness at the level of either noticing or understanding, and learners who reported in their think-alouds or questionnaires that they had used non-animacy based strategies (e.g., singular vs. plural), the quantitative results from the multiple-choice and production assessment tasks showed that unaware learners did not demonstrate any significant animacy bias in either the selection or the production of the trained or new determiner-noun phrases. In other words, their results were the opposite of those in Williams (2005).

Faretta-Stutenberg and Morgan-Short (2011), also conducted a replication of Williams (2005). Unlike Hama and Leow (2010), Faretta-Stutenberg and Morgan-Short strictly followed Williams’ study design, but like Hama and Leow, they used a finer-grained assessment of awareness which included the levels of noticing, understanding, and no report of awareness. The descriptive statistics showed that learners with awareness at the level of understanding were generally more accurate than learners with awareness at the level of noticing, and learners with awareness at the level of noticing were in turn more accurate than learners with no report of awareness. On generalization items, the Understanding group’s accuracy was significantly above chance; however, participants in the Noticing group and No Report group did not perform significantly
above chance. Like Hama and Leow (2010), this study also suggested that learners who lacked awareness did not show learning, in contrast to Williams’ (2005) findings.

Comparison of previous studies. In comparing the studies which found evidence supporting unaware learning and those which did not find such evidence, two methodological differences are apparent. First, the criteria used in Williams (2004, 2005) to classify learners as “aware” appear to correspond with what Schmidt called “understanding” or a higher level of awareness. In Williams’ (2004, 2005) studies, all of the aware learners were able to state the rule involving the animacy features of the determiners and nouns. As Schmidt has repeatedly argued, awareness at the level of understanding is facilitative but not necessary for learning. In other words, learning can occur without understanding. What is important to consider in studies that address unaware learning is noticing or a lower level of awareness. In Williams’ studies, it is quite possible to imagine that learners who had a lower level of awareness were included in the unaware learner group, thus affecting his results and leading him to interpret the data as showing evidence of unaware learning when this may have been unwarranted.

Second, the three studies reporting evidence of unaware learning used offline interviews as their only measure of awareness. As mentioned earlier, offline verbalization as a measure of awareness may not accurately represent learners’ actual cognitive processes. If online data cannot be collected due to the design of the studies, it is crucial 1) to provide some additional support for the retrieval of information, such as stimulus cues during the use of offline measures; 2) to ensure the minimization of the fabrication of reports by reminding learners to report only their original thoughts; and 3) to use another
awareness measure, or multiple measures, to triangulate the data. When all of these steps are not taken, studies suffer consequences, including 1) learners’ reports might be incomplete or inaccurate, or 2) at the time of the interview, they may report what they think they were thinking during the task, and this may not reflect their actual original thoughts. As Hama and Leow’s study (2010) showed, learners’ awareness of target rules may not be captured by one awareness measure; therefore, it is essential to triangulate data when investigating the issue of awareness.

Summary on Noticing

The studies with more sound methodologies as far as the measurement of awareness is concerned (Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leow, 2000) consistently suggested that noticing is necessary for L2 learning, as Schmidt and Robinson claimed (e.g., Robinson, 2003; Schmidt, 2001); however, further examination is undoubtedly essential in order to draw any firm conclusions. Collecting accurate data on noticing is, to say the least, a difficult task, and with the currently available awareness measures, noticing cannot be recorded with complete accuracy (Truscott, 1998); the zero-point claim of no learning without awareness is methodologically impossible to verify.

One generally acknowledged claim regarding awareness and learning in the L2 field is that awareness supports learning. The results from studies, especially the ones which utilized either think-alouds or stimulated recalls (oftentimes with an offline questionnaire or interview as an additional measure), have overwhelmingly suggested that awareness has facilitative effects on L2 learning, and that the more awareness they
show, the more learning they demonstrated as well (e.g., Alanen, 1995; Leow, 1997, 2000, 2001ab; Mackey, 2006; Rosa & Leow, 2004; Rosa & O’Neill, 1999; Sachs & Polio, 2007; Sachs & Suh, 2007; however, see Bell & Collins, 2009). The results from those studies emphasize the fact that awareness, regardless of its necessity, provides a very strong association with SLA.

In the following section, I will describe theoretical arguments and previous study findings regarding recasts, a form of interactional feedback. Many interactional feedback studies are based on the premise that noticing is at least important, if not necessary, for L2 development.

Recasts

Theoretical Arguments

Following the claim regarding the important role of noticing in SLA, the Interaction Approach argues that the negative evidence learners receive on their production through interactional feedback facilitates learners’ noticing and, thus, their L2 development. Gass (2003) has stated that while positive evidence is readily available outside of interaction, negative evidence is more difficult to obtain by other means. Negative evidence, which feedback can provide, is also claimed to be necessary to recognize certain types of errors that cannot be discovered solely though positive evidence (White, 1987). Below, an example of interactional feedback is provided (the example is taken from Mackey et al., 2000):
Example 1. Interactional feedback.

NNS: There is a basen of flowers on the bookshelf

NS: a basin?

NNS: base

NS: a base?

NNS: a base

NS: oh a vase

NNS: vase

In this example, a non-native speaker (NNS) used the non-nativelike lexical item, “basen”. This triggered the native speaker (NS) to provide interactional feedback, a clarification request, “basin?”. The non-native speaker realized the word “basen” was not correct and reformulated her utterance. Because the native speaker still did not understand what the non-native speaker was trying to say, she provided another clarification request, “a base?” This interactional episode ended with the native speaker’s recast, “oh a vase”, which the non-native speaker successfully incorporated into her subsequent utterance. Researchers such as Long (1996) claim that interactional feedback can serve to direct a learner’s selective attention to problematic aspects of his or her knowledge and production of an L2, possibly leading the learner to notice that what he or she says or knows differs from what a native speaker says. Schmidt and Frota (1986) called this process noticing the gap and proposed that it was an initial step in L2 learning. More recently, Gass and Mackey (2007) noted, “Feedback may help to make problematic
aspects of learners’ interlanguage salient and may give them additional opportunities to focus on their production or comprehension, thus promoting L2 development” (p. 182).

Recasts have been found to be the most frequently used type of interactional feedback in instructional settings (e.g., Ellis, Basturkmen, & Loewen, 2001; Lyster & Mori, 2006; Lyster & Ranta, 1997). Gass and Mackey (2007) define a recast as “a rephrasing of an incorrect utterance using a correct form while maintaining the original meaning” (p. 182) (see Example 1). One of recasts’ unique and helpful characteristic is that recasts reformulate learners’ utterances while keeping their central meanings intact; thus, the learners understand all or at least part of the recasts their interlocutors provide. This frees learners’ attentional resources from focusing on meaning, and in turn, the resources may be allocated to the linguistic form of the more targetlike version (Long, 1996, 2007). Secondly, these correct forms are provided immediately after learners’ erroneous utterances; thus, recasts juxtapose deviant learner output with targetlike formulations. According to Long (2007), “This allows the learner to compare the two forms side by side, so to speak, and to observe the contrast” (also see Long, 1996; Saxton, 1997). Thirdly, in contrast to other types of interactional feedback such as negotiations, recasts not only provide negative evidence to alert learners of a mistake, but also supply positive evidence. Therefore, similarly to other types of interactional feedback, recasts can provide conditions which help learners notice the gap through their negative evidence, and at the same time, they also generate opportunities for learners to notice new information via their positive evidence. In other words, recasts can directly assist learners to create new knowledge (Long, 2007).
Recasts and L2 Development: Empirical Evidence

Since the mid-1990s, the link between recasts and L2 development has been extensively tested, and the results from those studies have largely demonstrated that recasts have a positive effect on learning (See a meta-analysis from Mackey & Goo, 2007). With this firm support from previous empirical evidence, recent interaction research has not focused on whether recasts impact learning but on how they assist learning and under what conditions they are more likely to bring about SLA (e.g., Mackey, 2012; Mackey et al., 2012).

Researchers, so far, have found empirical evidence that several factors influence the efficacy of recasts on L2 development.¹ These include, but are not limited to: 1) the type of recast (e.g., Loewen & Philp, 2006; Nassaji, 2009), 2) the linguistic target (e.g., Iwashita, 2003; Leeman, 2003; Mackey, 2006; Ortega & Long, 1997; Yang & Lyster, 2010; Yilmaz & Yuksel, 2011; Witzel & Ono, 2003), 3) developmental readiness (e.g., Mackey & Philp, 1998) or proficiency level (Ammar & Spada, 2006), 4) individual differences (e.g., Goo, 2011; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey & Sachs, 2012; Sagarra, 2007; Sheen, 2008; Trofimovich, Ammar, & Gatbonton, 2007), and 5) task variables (Révész, 2009; Révész & Han, 2006; Romanova, 2010). Below, I will briefly describe claims which have been made for each factor and the empirical findings from previous studies (see also Mackey, 2012).

¹ The results from studies utilizing written recasts are not included here. This is indeed one of the areas that interaction researchers should investigate, and their results have provided and will provide significant contributions to interaction research. However, I address only oral recasts here due to the facts that 1) some studies (e.g., Lai & Zhao, 2006; Yilmaz & Yuksel, 2011) have shown written recasts may be different from oral recasts, and 2) my dissertation incorporated oral recasts.
Types of recasts. Recasts are considered by some, categorically, to be an implicit type of corrective feedback (e.g., Long, 1996; Long & Robinson, 1998) which does not overtly indicate to learners that they have made a mistake. However, it is widely acknowledged that the degree of implicitness among different types of recasts varies greatly (e.g., Ellis & Sheen, 2006; Nicholas, Lightbown, & Spada, 2001), and that certain types of recasts may be more salient than others, with some recasts’ corrective intentions appearing quite obvious to learners. This has made researchers argue that the efficacy of recasts on learning may change depending on the type of recasts learners receive (e.g., Doughty & Varela, 1998; Han, 2002).

Empirically, this account regarding the differentiated efficacy of varied types of recasts on development has been supported. Loewen and Philp (2006) revealed that characteristics of recasts such as having a lower number of morphemes, a single change, and/or interrogative intonation were associated with accuracy on posttests. In more recent study, Nassaji (2009) found higher accuracy on immediate and delayed posttests for isolated recasts (i.e., recasts which segment learners’ erroneous utterances and include only the problematic parts) compared to recasts which are not segmented, and also for recasts with interrogative intonation and/or added stress compared to recasts without those two characteristics.

These results overall are consistent with the prediction that recasts which are arguably more salient and explicit should lead to L2 development. However, regarding the intonation of recasts, Loewen and Philp (2006) predicted that declarative intonation
should facilitate development more effectively compared to recasts with interrogative intonation because,

an interrogative recast is ambiguous as corrective feedback because the learner can interpret it either as corrective or as a request to confirm the intended meaning…declarative recasts being arguably more explicit as corrective feedback than interrogative recasts because they do not suggest a choice for the learner to accept or refute (pp. 540-541).

Therefore, the results from Loewen and Philp (2006) and Nassaji (2009) contradicted expectations. In order to explain their contradicting findings, Loewen and Philp alternatively speculated that “declarative recasts may be interpreted in the same way as a teacher’s repetition of learners’ TL [targetlike] forms (see Lyster, 1998) and, thus, may function just as implicitly for the learners” (p. 541). Taking into account this explanation from Loewen and Philp, it may be that the intonation of recasts itself does not change their degree of saliency and explicitness. The intonation of recasts and their efficacy will be addressed again in a later section on recasts and learners’ noticing.

Target of recasts. In a very early interaction study of two Vietnamese children learning English, Sato (1986, 1990) mentioned that a positive link between interaction and development may not be found for certain types of structures. This argument has been upheld by the results from a number of recast studies (e.g., Iwashita, 2003; Leeman, 2003; Mackey, 2006; Ortega & Long, 1997; Yang & Lyster, 2010; Yilmaz & Yuksel, 2011; Witzel & Ono, 2003). These studies examined the effects of recasts on multiple language features (e.g., number agreement, adverb placement, past tense) and found that
learners benefited differentially from recasts depending on the features the recasts targeted. The researchers of the abovementioned studies all inferred that this varied efficacy was most likely due to differences in the saliency of those features; that is, when the target of a recast is perceptually more salient, the recast may become more noticeable and thus effective.

Considering the saliency of linguistic target as a plausible reason for the differential efficacy of recasts, a few studies have investigated this very connection (Witzel & Ono, 2003; Yilmaz & Yuksel, 2011). Inspired by Goldschneider and DeKeyser (2001), the first study to undertake this topic, Witzel and Ono (2003), produced mixed results. They incorporated four forms: plural –s, past –ed, third-person possessive, and progressive –ing. On the basis of phonetic substance, syllabicity, sonority, and bound/free status, the first two forms were categorized as non-salient and the latter two as salient. They found recasts to be effective for the salient forms (third-person possessive, progressive –ing) and not for one of the non-salient forms (plural –s), as expected. However, on the other non-salient form, past –ed, they found a positive impact of recasts, just as with the salient forms. The results from another study (Yilmaz & Yuksel, 2011) indicated no difference in the efficacy of recasts on the salient form, the Turkish plural morpheme, versus the non-salient form, a Turkish locative case morpheme. Yilmaz and Yuksel explained this result by pointing out that they had determined the two forms’ perceptual saliency according to their phonetic substance, syllabicity, sonority, and suffix length; however, the non-salient form was sentence final whereas the salient form was
not. Therefore, considering the well attested salience of edge positions, this may have created a smaller salience gap between the two types of forms.

The operationalization of salience is a difficult task because many factors must be considered (Ellis & Sheen, 2006; Long, 2007). Beyond the sorts of perceptual features mentioned above, salience can be influenced by learners’ language backgrounds and previous L2 experiences. It is also possible to conjecture that the mixed results from Witzel and Ono (2003) and Yilmaz and Yuksel (2011) may have been caused by factors that are not related to the target of recasts, but to learner factors such as development readiness and memory. In fact, the learners in Yilmaz and Yuksel’s study had no knowledge of Turkish prior to the experiment; therefore, it is possible that they were not ready to learn the target forms. In addition, Yilmaz and Yuksel’s non-salient form, which appeared at the end of the sentence, likely had fewer constraints related to learners’ memory capacities, and this may have washed out any effect related to saliency.

**Developmental readiness and proficiency.** According to Pienemann’s Teachability Hypothesis (1984, 1989), L2 learners are constrained to a natural developmental sequence of acquisition; therefore, pedagogical intervention can speed up the rate of development following the sequence, but it cannot alter or allow learners to skip the stages in the sequence. This hypothesis claims that instruction should be one stage beyond the learners’ current developmental level, and that any pedagogical interventions well beyond learners’ current stages are likely to be unsuccessful. In the area of interaction research, Mackey and Philp (1998) have provided empirical support for the differential efficacy of recasts on learners who were either developmentally ready
or unready to learn the linguistic target, question forms. Using the developmental sequence for question formation identified by Pienemann and Johnston (1987), Mackey and Philp (1998) determined learner readiness and analyzed the extent of their learning after interaction with recasts. They found that while the ready learners showed improvement in the production of question forms, the unready learners did not.

Similar results were reported in Ammar and Spada (2006). They examined the efficacy of recasts on learning English third-person possessive determiners, a difficult aspect of English grammar for Francophone learners of English, for two different proficiency groups. They found that the high-proficiency learners seemed to benefit from recasts, but this was not the case for the low-proficiency learners. The results from these studies indicate that learners’ current knowledge in L2 is associated with the amount of benefit gained from recasts.

**Individual differences.** Long (1996), in his seminal article on interaction and L2 learning, stated that “success or failure to learn can rarely if ever be attributed to the environment alone. Part of the explanation lies inside the learner, most importantly in the areas of attention, awareness, and cognitive processing” (p. 425). Researchers such as Robinson (1995) and Schmidt (2001) argued that cognitive factors influence learning.

One cognitive factor that has been receiving a fair amount of attention in SLA is working memory (WM). WM is “an integrated system for temporarily storing and manipulating information” (Baddeley, 2003, p. 837). It underlies learners’ ability to handle both processing and storing of information simultaneously as they work on a language task; therefore, it has been claimed that learners with higher WM may receive...
more benefit from recasts. A number of researchers have empirically tested this claim (Goo, 2011; Mackey et al., 2002; Mackey & Sachs, 2012; Révész, 2012; Sagarra, 2007; Trofimovich et al., 2007) and have revealed somewhat mixed results.

Révész (2012), Sagarra (2007), and Mackey and Sachs (2012) found the expected association between WM and the efficacy of recasts. In Révész (2012), each learner engaged in one-on-one communicative tasks with the researcher. When learners committed errors on past progressive forms, the researcher provided declarative recasts. The results demonstrated that WM was related to the extent of development achieved by the learners through treatments with recasts. In Sagarra (2007), learners engaged in a computerized task on Spanish noun-adjective agreement. Throughout the task, they received computer-delivered recasts. The results showed that WM was related to the learners’ subsequent L2 performance on both immediate and delayed posttests. In Mackey and Sachs (2012) participants, age 65-89, engaged in dyadic communicative tasks and received recasts. The results revealed that only learners with higher WM capacity showed development in the immediate posttest, and learners with the highest WM showed sustained development at their one-month delayed posttest.

Other studies, however, did not show such a clear pattern. Mackey et al. (2002) found a positive relationship between WM and L2 performance on the delayed posttests, but on the immediate posttests it was the learners with lower WM who showed improvement. In their study, each learner was paired with a native speaker to carry out communicative tasks designed to elicit English question forms. When the learners’ questions were nontargetlike, they received recasts. On the immediate posttest, learners
with lower WM showed improvement on question forms; however, they did not show sustained improvement on a two-month delayed posttest. In contrast, learners with higher WM did not demonstrate development during the immediate posttest, but development was found during the delayed posttest.

Contradicting results from the above studies were reported in Trofimovich et al. (2007). In their study, learners participated in a computerized task where they described pictures and received feedback (i.e., a grammatically correct way of describing the picture) irrespective of the existence of an error in their utterance. The results did not show any significant association between WM and learners’ production accuracy on posttests. Trofimovich et al. argued that due to the design of their study, learners received more consistent and frequent input regarding the target forms compared to other recast studies, possibly diminishing the influence of WM capacity.

All in all, empirical studies have not produced consistent results regarding WM. However, the few studies which have been conducted in these areas so far have indicated that such relationships exist. This topic, therefore, remains potentially important when we consider the efficacy of recasts on development (see Mackey, 2012).

Much of the research into relationships between individual differences and interaction-driven learning has focused on learners’ cognitive factors such as WM. However, it has also been found that the psychological or affective factor of anxiety may impact the L2 developmental benefits learners receive from recasts on L2 development. In a study by Sheen (2008), learners participated in two narrative retelling tasks as a whole-class activity. As students took turns retelling the stories a few lines at a time, the
teachers provided recasts on erroneous uses of English articles. The researcher then divided the students into two groups, high-anxiety learners and low-anxiety learners, based on their responses on a questionnaire. The results showed that the low-anxiety recast group outperformed the high-anxiety recast group on both immediate and delayed posttests. This study underscores that learners’ psychological or affective factors, which are often not considered in the empirical studies, deserve future examinations.

**Task variables.** The nature of the tasks which are employed to support conversational interaction in a second language has always been a research focus of the interaction approach. In the early stages of interaction research, there was extensive examination of what types of tasks led to more interactional modifications because these were theorized to provide opportunities for L2 development (see Pica, Kanagy, & Falodun, 1993). The relationship between task variables and learners’ L2 performance has also been extensively examined in the literature on task-based learning using two models of attention, a single-resource model and a multiple-resource model, both which assert that the cognitive demands of tasks can change the way learners’ attentional resources are deployed during task performance (Robinson, 2001b, 2003; Skehan, 1998; Skehan & Foster, 2001).

Recently, researchers began connecting the interaction approach and task-based learning research. It has been argued that because task variables can affect the extent of learner attention to form during L2 production, they may also affect the efficacy of recasts (e.g., Long, 2007; Nicholas et al., 2001; Robinson, 2003, 2005; Robinson & Gilabert, 2007; Skehan & Foster, 2001). A few studies have been conducted to examine
this claim empirically (Révész, 2009; Révész & Han, 2006; Romanova, 2010\(^2\)) and indicate that task variables are undeniably a factor which may affect the benefits learners obtain from recasts.

The first empirical study to tackle this topic was Révész and Han (2006). They examined whether task content familiarity and task type would influence recasts’ efficacy. Participants were divided into two groups which differed in terms of task content familiarity: one with a task containing the same content in all treatment sessions and the other with a task containing different content in each of the treatment sessions. Each of those groups was further divided into two groups according to task type: one with a video-primed task and the other with a note-primed task. During the three treatment sessions, after viewing either the video prime or the note prime, the learners in those four experimental groups engaged in event-retelling activities (i.e., describing the event shown in the prime) and received recasts on erroneous uses of past progressive forms. The results from written posttests did not show a clear advantage for any group. However, the results from oral immediate- and delayed-posttests showed that 1) learners who repeatedly had the same content during the treatments significantly outperformed learners who completed tasks with different content, and 2) learners who completed the video-primed tasks significantly outperformed learners who completed the note-primed task.

In another study by Révész (2009), she examined the provision of contextual support in relation to the benefits of recasts. During the treatment, first, all learners were presented with ten photos describing an event, and then they retold the events. A group of

\(^2\) Romanova’s study is not described in this section, Recasts. See the later section, Strategic Planning, Noticing, and L2 Development, for its description.
students who received recasts on their errors regarding past progressive forms during the event-retelling treatment activity were divided into two groups. One of the groups was asked to describe the event with the presence of those photos as contextual support. The other groups did not have those photos. The results from immediate and delayed posttests showed that the learners without photos (i.e., without contextual support) outperformed the learners with photos.

The results from the studies above indicate that task variables do, indeed, affect the benefits of recasts. However, more research is unarguably warranted to understand how different task variables interact with the efficacy of recasts. In the later section, Strategic Planning, Noticing, and L2 Development, I will address the variable most tested in task-based learning literature, strategic planning, and its effect on the efficacy of recasts.

Recasts and Noticing: Empirical Evidence

Upon witnessing the consistently positive findings regarding the impact of recasts on L2 development, interaction researchers have begun seeking to identify the cognitive processes involved in interaction and language development. This centers their focus on noticing and recasts, that is, they have started exploring one of the theoretical premises of recasts--the theory that recasts may lead learners to notice their utterances are nontargetlike and/or specific nontargetlike aspects (e.g., Philp, 2009). Examinations of

---

3 Here I refrained from explaining how the two models of attention, a single-resource model and a multiple-resource model, explain the results of the two studies by Révész. This is because the explanation would require significant amount of space and the focus here is to simply show that task variables do alter the efficacy of recasts. I will describe the two models of attention in the later section, Task Planning and Task Performance. How the models implications for noticing and efficacy of recasts will be explained in the last section of this chapter, Strategic Planning, Noticing, and L2 Development.
learners’ noticing when they were given recasts have been undertaken using introspective measures such as stimulated recall (Egi, 2007a; Gass & Lewis, 2007; Kim & Han, 2007; Mackey, 2006; Mackey, Al-Khalil, Atanassova, Hama, Logan-Terry, & Nakatsukasa, 2007; Mackey et al., 2002; Nabei & Swain, 2002; Roberts, 1995), cued immediate recall (Philp, 2003), immediate report (Egi, 2007a), and private speech (Ohta, 2000). These studies have revealed that learners do, indeed, notice the language in recasts. Moreover, many of the factors mentioned above in relation to L2 development (e.g., type of recast, linguistic target, individual differences) have been found to influence noticing in a similar manner. The complementary patterns found in both development and noticing studies provide support, albeit indirect, for the proposal that those factors affect development because they change what they notice. Consistent with this idea, Mackey (2006) found that some language features elicited more noticing than others and that learners showed the most development on those most noticed features.

Types of recasts. Corroborating the results from developmental studies, examinations relating noticing and types of recasts have found that 1) segmented recasts (i.e., isolated recasts) yield more learner noticing compared to those which are non-segmented (Kim & Han, 2007; Roberts, 1995), and 2) recasts with fewer morphemes and a lower number of corrections have a stronger relationship with reported noticing compared to those with larger numbers of morphemes and corrections (Egi, 2007a; Philp, 2003). These results suggest that recasts which are arguably more salient and/or explicit are associated with more noticing. The results in regards to the intonation of recasts (i.e., declarative vs. interrogative) have depicted a mixed pattern; specifically, more noticing
was found with interrogative recasts in one study (Mackey et al., 2007), but the opposite findings were revealed in the other (Kim & Han, 2007). This contradicting pattern supports the conclusion drawn from Loewen and Philp’s (2006) developmental study, based on which the researchers stated that the intonation of recasts alone cannot determine whether recasts are more salient and/or explicit or whether they will therefore lead to learners’ noticing and development.

**Target of recasts.** Mackey (2006) found, similarly to the development studies, that learners’ noticing can differ according to the language features which are being targeted by recasts. Her results showed that noticing occurred more often when recasts targeted question forms compared to past tense—perhaps, she explained, due to the difference in the saliency of those forms. Whereas language features have been tested mainly in developmental studies, general areas of language (e.g., morphology, lexis) have also been investigated in introspective studies (Egi, 2007a; Gass & Lewis, 2007; Kim & Han, 2007; Mackey et al., 2007). These studies generally showed that lexis yields the most noticing, followed by morphology, and finally syntax. This pattern, Mackey et al. (2000) explained, may also be due to the saliency of those areas of language. Kim and Han (2007) alternatively stated that the meaningfulness of the area of language for communication may have been a reason for the pattern to emerge.

**Proficiency.** As in the development studies, learner factors such as proficiency also influence the noticing of recasts. Philp (2003) assigned her participants into high-, intermediate- and low-level groups based on their production of question forms according to the six stages of development for question formation (Pienemann & Johnston, 1987).
Philp operationalized noticing as the accurate immediate recall of recasts. During a picture drawing task, learners asked a native interlocutor questions to complete the picture drawing and the native interlocutor provided recasts to any nontargetlike utterances, particularly question forms. Following each recast, the native interlocutor knocked on the table to cue the learner to recall the recast she received. The Philp found that learners with higher proficiency were able to recall recasts more accurately compared to learners with lower proficiency, complementing the results from development/proficiency studies which revealed that learners who are developmentally ready or have higher proficiency are more likely to benefit from recasts.

**Individual differences.** In studies of interaction and L2 development, WM is an individual difference factor which has produced rather mixed results. Regarding the relationship between WM and the noticing of recasts, Mackey et al. (2002) found marginally significant results. In their study, learners were classified as high, middle, or low in terms of their WM scores and as high or low with regard to the amount of noticing they reported. It was found that six out of the seven learners who reported more noticing were in the high WM group and one was in the low WM group. Of the 13 learners who reported less noticing, four were in the high group and nine were in the low group. This appeared to present a clear positive relationship between WM and noticing; however, a statistical analysis showed that working memory was only marginally related to the amount of noticing learners reported.
Summary on Recasts

Research thus far has shown that there are multiple factors which appear to affect the efficacy of recasts on learners’ development and noticing. In addition, the results are consistent with the idea that recasts facilitate noticing and thereby support development. Under what conditions recasts are more effective is still a question remaining to be fully answered. Among the factors discussed here, task variables, which have been a focus of interaction research since its inception, are now coming back into focus with a new research agenda. Another strand of SLA research, task-based language learning, has extensively examined the impact of task variables on learners’ immediate language production. This primary focus has been different from the focus on L2 development found in recent work on recasts. However, more recent studies in task-based language learning have indicated task variables to be an important fact to consider in recast research (Révész, 2009; Révész & Han, 2006; Romanova, 2010). In the next section, I will address the task variable, task planning, the most tested variable in task-based learning literature.

Task Planning and Task Performance

Types of Task Planning

In two review papers on task planning research, Ellis (2005, 2009) identifies three types of planning: rehearsal, strategic planning, and within-task planning. These can be categorized into two principal types of planning: pre-task and within-task. The ostensible difference between the two types of planning is the time at which it takes place. Pre-task planning, as the name indicates, occurs prior to task performance. Pre-task planning can
be further categorized as involving 1) rehearsal and/or 2) strategic planning. During rehearsal, learners are provided an opportunity to perform the task prior to the actual task performance. It therefore entails repetition of performance. During strategic planning, previous studies have provided learners with 1 to 15 minutes of planning time. Participants receive the materials and instructions for an upcoming task then prepare for the performance by considering how they want to perform the task. In contrast to pre-task planning, within-task planning occurs during the performance of a task. Yuan and Ellis (2003) theorized it as “involving a particular kind of speech production that incorporates both ‘careful’ production (as opposed to ‘rapid’ production) and ‘monitoring’” (p. 5). Within-task planning, therefore, can be further differentiated according to whether or not task performance is pressured or unpressured. When students are allowed more online performance time, they are assumed to engage in covert planning activities more often than students who are put under time pressure. Unlike the two aforementioned types of pre-task planning (i.e., rehearsal and strategic planning), the difference between pressured and unpressured task performance is not a dichotomous distinction; they are on a continuum.

**Measures of Task Performance**

Among the three types of task planning introduced above, strategic planning has been the main focus of investigations in task planning research. Strategic planning studies, as well as other types of planning studies, have thus far been conducted chiefly in task-based learning frameworks, which have tended to focus on how task-related variables impact learners’ production during task performance. In general, learners’ production is

Fluency refers to the ability to convey meaning in real time; therefore, higher levels of fluency requires learners to access and deploy ready-made chunks of language. Complexity and accuracy, on the other hand, are both concerned with language form and syntactic processing. Complexity refers to the ability to use more advanced language. It may imply learners’ willingness to take risks using language structures that they still struggle to control effectively; therefore, it has been argued that complexity may be related to restructuring or changes in learners’ current L2 grammar systems. Accuracy refers to the ability to avoid errors. It may involve learners’ ability to control their existing language resources or their avoidance of advanced language which may provoke errors.

Strategic planning literature on task planning argues that these aspects of L2 performance can be modulated by the implementation of strategic planning. Most importantly, it has been claimed that strategic planning can encourage learners to focus on complexity and accuracy, thereby promoting syntactic processing.

**Strategic Planning**

*Two theoretical models.* Two theoretical accounts of attention allocation during task completion, the single-resource model and the multiple-resource model, despite making different assumptions about the nature of attention, provide similar predictions as to how strategic planning may affect learners’ L2 production during task performance. The single-resource model (Skehan, 1998; Skehan & Foster, 2001; VanPatten, 1996)
argues that humans have limited attention capacity; thus, when the content or meaning of a task demands a significant amount of attention, learners respond to the pressure by focusing on meaning to complete the task, leaving an insufficient amount of attention available to allocate to its language. Therefore, reducing the attentional demand for content or meaning is argued to free attentional resources which can then be used to focus on more formal aspects of the L2 code. The application of strategic planning prior to task performance can reduce the attentional demands involved in completing a task and may therefore increase the chance that learners will attend to form while engaged in meaning-making. This means that strategic planning may have a positive effect on learners’ production--not only in their fluency but also on the form-oriented aspects of production, accuracy and complexity.

The multiple-resource model (e.g., Robinson, 2007; Wickens, 1989, 1992) proposes that attention can be drawn from multiple sources that are independent of each other. Thus, the attention demanded by the performance of a task need not necessarily discourage learners from attending to language. On the contrary, in some cases, a cognitively demanding task can lead learners to pay more attention to language compared to a less cognitively demanding task. In Robinson’s (e.g., 2001, 2003, 2005) Cognition Hypothesis, which was built on the multiple-resource model, he argued that features of tasks and the ways tasks are implemented can be manipulated to increase or lessen the cognitive demands that tasks impose on learners. Such manipulations can occur in two sets of task variables: resource-directing dimensions, which direct learners’ attentional and memory resources to aspects of the L2 system, and resource-dispersing dimensions,
which do not direct learners to any aspect of the linguistic system. When cognitive demands are increased in resource-directing dimensions, it is predicted that learners’ attention to relevant feature of language is facilitated. However, when cognitive demands are increased in resource-dispersing dimensions, the amount of attention and memory resources available for language-related concerns is reduced. Robinson argued that because task planning lies within the area of resource-dispersing factors when learners are not given time to plan prior to completing a task, their L2 performance is expected to be affected negatively in all aspects, fluency, accuracy, and complexity. Even though Robinson’s model does not specifically state that strategic planning should facilitate learners’ attention to language as Skehan predicts, it does explain that the provision of strategic planning may help avoid negative consequences to learner production.

*Levelt’s model of speech production.* More fine-grained insights into how strategic planning may affect following production can be drawn from Level’s model of speech production. Levelt posits three autonomous components of the speech production system: a Conceptualizer, a Formulator, and an Articulator. Levelt describes each component as being responsible for different aspects of speech production. The Conceptualizer, the first production processing component, is in charge of planning the conceptual and pragmatic content of speech. In this stage, speakers determine the overall communicative goal, develop series of sub-goals based on the overall goal, then gather the information needed to realize the sub-goals. The outcome of this process is a preverbal message; i.e., a message that is not linguistic in nature. The Formulator, the next processing component, receives the preverbal message from the Conceptualizer and
creates a linguistic representation of the message by drawing lexical items from the speaker’s mental lexicon. Once a lexical item has been activated, subcategorization properties and pronunciation are also activated. Finally, the internally formulated utterance in the Formulator is sent to the Articulator, the final component, and the message is given its phonetic shape and articulated. These processes are regulated by three monitoring systems. The first monitor checks whether the preverbal message matches with the speaker’s intentions prior to sending the information to the Formulator. The second monitor inspects the internal linguistic plan against the overall plan before the conversion into overt speech takes place. Lastly, the articulated speech is inspected by the third monitor.

Considering Levelt’s speech production components, Ellis (2005) has hypothesized the possible outcome of strategic planning:

> Strategic planning can be considered likely to assist conceptualization in particular and thus contribute to greater message complexity, and also to enhance fluency (p.14)

Regarding accuracy, Ellis, however, did not predict that strategic planning would be helpful; rather, he argued that within-task planning, which may provide learners needed time for the controlled processing required for monitoring, would be beneficial.

*Empirical studies.* As predicted by the two models, studies have shown that the implementation of strategic planning generally produces positive effects on learners’ task performance. The studies have revealed clear and consistent confirmatory effects on fluency (e.g. Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2006; Kawauchi, 2005;
Mehnert, 1998; Ortega, 1999; Sangarun, 2005; Skehan & Foster, 1997, 2005). For complexity and accuracy, with the exception of a few studies, research has also reported encouraging results (e.g., complexity: Crookes, 1989; Foster & Skehan, 1996, 1999; Gilabert, 2006; Kawauchi, 2005; Mochizuki & Ortega, 2008; Ortega, 1999; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 1997; Yuan & Ellis, 2003; accuracy: Crookes, 1989; Foster & Skehan, 1996, 1999; Kawauchi, 2005; Mehnert, 1998; Mochizuki & Ortega, 2008; Ortega, 1999; Sangarun, 2005; Skehan & Foster, 1997, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 1997). However, many of the beneficial effects found for strategic planning on complexity and accuracy have been selective. In other words, depending on certain factors, the expected results may or may not be apparent. These factors include but are not limited to: 1) amount of planning time, 2) linguistic features, 3) learner proficiency, 4) type of tasks used, 5) length of task performance, 6) availability of within-task planning, and 7) planning instructions.

Amount of planning time. Mehnert (1998) compared learners’ task performances following different amounts of strategic planning time (0, 1, 5, or 10 minutes). She found that for one of the tasks used, the L2 production of the learners who had only 1 minute of planning time was significantly more accurate compared to the L2 production of learners without planning time, but that no advantage was found for learners with more planning time, whether 5 or 10 minutes, over learners without planning. In contrast, for complexity, even though statistical significance was not achieved, it was found that learners with 10 minutes of planning time produced the most complex language compared to any other group. Mehnert interpreted the results to mean, perhaps, that learners differently prioritize
what to plan depending on the amount of planning time allotted. In other words, learners may focus on accuracy when given limited planning time, and complexity when given a greater amount of time.

Linguistic features. While the majority of the strategic planning literature employed global measures of learners’ production accuracy (e.g., error-free clauses), arguing that this would yield a more realistic picture of learners’ performance (e.g., Foster & Skehan, 1996; Skehan & Foster, 1997), some have focused on the accuracy of specific forms in learners’ production (e.g., Crookes, 1989; Ellis, 1987; Kawauchi, 2005; Mochizuki & Ortega, 2008; Ortega, 1999, 2005; Wigglesworth, 1997). Ortega (1999) has claimed, “global measures have the disadvantage of being too broad to capture small changes…and obscure errors in grammatical domains that may be important at a given level of development” (p. 118). In the studies that have utilized specifically targeted measures, strategic planning has been found to facilitate the accuracy of linguistic forms selectively. For example, in a study by Ellis (1987) which examined different types of past tense morphemes, learners showed significantly greater accuracy with regular past tense forms in planned productions compared to unplanned, but hardly any difference in accuracy with irregular past tense forms in the two types of productions. Ellis argued that this may have occurred because regular past tense forms (e.g., -ed) are learned forms which can be applied to planned discourse, but when there is no time to plan, it can become quite difficult to monitor and maintain accurate use of productive morphology throughout a task. Irregular past tense forms, on the other hand, occur frequently in L2 input and may be highly automatized; therefore, they may not be very sensitive to
planning. Kawauchi (2005) also examined the use of past tense morphemes (i.e., copula, regulars and irregulars) among learners at three proficiency levels and demonstrated the impact of strategic planning interacted not only with different types of past tense morphemes but also learners’ proficiency. This factor will be discussed in greater detail in the following.

*Proficiency.* Proficiency has been examined by a number of researchers (e.g., Kawauchi, 2005; Ortega, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 1997). Wigglesworth (1997) reported that among higher-proficiency learners, those with strategic planning time produced significantly more accurate and more complex language in one of the difficult tasks than those without strategic planning time; however, among lower-proficiency learners, no benefit of planning was found in any task (for less difficult tasks, there was no difference between the two group of learners). Ortega (2005) similarly found that advanced-level learners significantly benefited from strategic planning in terms of linguistic accuracy, but such benefits were not apparent in low-intermediate learners. Ortega (2005), who investigated the strategies learners utilized during strategic planning time summarized her findings:

[A]mong advanced level speakers, it [planning time] allowed a more balanced commitment of effort to retrieval and rehearsal and a fuller engagement with self-monitoring strategies, while among low-intermediate speakers it fostered retrieval strategies committed to solving lexical and verbal morphology problems (p. 105).

The patterns from these two studies appear to indicate that strategic planning may be more useful for learners with higher proficiency compared to lower proficiency, and that
strategic planning may be used differently, and perhaps more effectively, by learners with higher proficiency.

A glance at another study, Kawauchi (2005), which examined multiple proficiency levels (low, high, and advanced), however, suggest contradictory results to those found by Ortega and Wigglesworth. For complexity, the benefit of strategic planning was most apparent with learners of high proficiency. For accuracy, lower-proficiency learners actually showed the most improvement following strategic planning. Kawauchi explained the results by suggesting that, in relation to complexity, learners of high proficiency may have benefited most from the planning because they had adequate L2 knowledge to draw from during planning, and the planning enabled them to access this knowledge during task performance. For the advanced-level learners, ceiling effects may have occurred. As for accuracy, it may again be the case that there was a ceiling effect for both high and advanced learners.

These above studies appear to indicate that one cannot simply draw the conclusion that higher-proficient learners demonstrate more improvement following strategic planning than lower-proficient learners, or vice-versa. It seems that the pattern is not straightforward, and that the tasks in which learners engage may influence the benefit learners at different proficiency levels gain from the planning. Ortega (1999) in fact found that 12 out of 32 advanced-level learners did not find the planning opportunities an advantage. Several participants stated that the stories used for her narrative tasks were simple enough and had a clear and inherent enough structure to render planning unnecessary. In such a case, obviously, not much benefit from strategic planning would
be found among the learners with higher proficiency. Overall, the findings from the studies above indicate that the effects of strategic planning time on task performance interact with proficiency, however, pattern may not be as obvious.

Type of tasks used. It has been argued that the usefulness of strategic planning may be greater when learners engage in cognitively demanding tasks (e.g., Foster & Skehan, 1996; Skehan & Foster, 1997). Based on this claim, several researchers carried out experiments designed to test the interaction between the effects of strategic planning and cognitive difficulty (Foster & Skehan, 1996; Mehnert, 1998; Skehan & Foster, 1997; Wigglesworth, 1997). Foster and Skehan (1996) and Skehan and Foster (1997) explored three task types: a personal information exchange task, a narrative task, and a decision-making task, which they argued had different levels of cognitive difficulty. The results from each of those two studies were mixed and did not appear to support the prediction. Furthermore, the results of the two studies contradicted each other. Mehnert (1998), which used two tasks differing in complexity, also failed to confirm the prediction. Mehnert used two tasks which required participants to leave a message on friend’s answering machine. The tasks differed in the cognitive difficulty: the instruction task, the easier task, which tapped familiar information, was structured, and used present tense or future tense; the exposition task, the more difficult task, was based on unfamiliar information, was unstructured, and used past tense. The results showed that the effect of strategic planning on the two tasks was not significant.

Wigglesworth (1997), the aforementioned strategic planning study examining proficiency and tasks with varying levels of cognitive difficulty, however, found some
evidence to support the argument. Based on Wigglesworth’s results, it is possible to conjecture that intervening factors, such as proficiency, may have obscured the connection between task complexity and the efficacy of strategic planning in the other studies. Ortega (1999) also raised a question regarding Foster and Skehan (1996) and Skehan and Foster’s (1997) hypothesized differences between task types (e.g., a narrative task, a decision-making task) by arguing that large differences may be found even within the same task type.

The previous studies appear to present several problems; however, Foster and Skehan (1996) and Skehan and Foster’s (1997) post-hoc interpretations of their results present some interesting insights regarding tasks and the efficacy of strategic planning on accuracy and complexity. A closer look at their data showed some characteristics of tasks to be related to accuracy but not complexity. That is, with tasks where gain in accuracy from strategic planning occurred, each contain a clear inherent structure. These clear macrostructures, they argued, do not lead learners to use complex languages to express ideas; therefore, their attention resources may be spared to attend to accuracy. This explanation, however, needs to be examined in future studies.

Length of task performance. Skehan and Foster (2005) have pointed out that the typical practice of task-based studies, in which only a brief (e.g., 5-minute) excerpt of the early stage of task performance is actually analyzed, deters researchers from seeing a full picture of how strategic planning influences task performance. In their 2005 study, they compared the first five minutes and the second five minutes of learners’ task performances following strategic planning and revealed a clear and rather discouraging
pattern. On both accuracy and complexity measures, for the first five minutes, learners who had time to plan with detailed instructions produced significantly more accurate and complex utterances compared to learners without planning time. However, no significant difference between planners and non-planners was found in any measure of accuracy or complexity for the second five minutes of performance. This difference in the level of performance between the first five minute and the second five minute for planners indicates, according to Skehan and Foster, that the sustainability of the effects of strategic planning may be limited.

_Availability of within-task planning._ Wendel (1997) and Yuan and Ellis (2003) argued that a key factor in whether learners produce accurate utterances depends on whether learners are given within-task planning, in other words, if they are given time to engage in careful online planning while they perform a task. This claim, thus far, had only been empirically tested by Yuan and Ellis. In their study, three planning conditions were compared: no planning with limited time for task performance, strategic planning with limited time for task performance, and within-task planning with unlimited time for task performance. They reported that learners in the within-task/unlimited-time group outperformed learners in the no-planning/limited-time group in all accuracy measures. However, no advantage in any accuracy measures was found for learners in the strategic planning/limited-time group compared to learners in the no-planning/limited-time group. As regarding fluency and complexity measures, there was a general trend that the strategic planning/limited-time group outperformed the within-task/unlimited-time group. The researchers claimed, based on notes taken by the learners during the strategic
planning time and comments from post-exposure questionnaires and interviews, that planning prior to performance may predispose learners to focus on content and its organization. This in turn positively affects fluency and complexity but not accuracy. In addition, learners in the no planning and strategic planning groups, those not allotted unlimited time for task performance, commented that it was difficult to monitor their performances because of the need to focus on the ongoing meaning-making activity. Yuan and Ellis’s study appear to show that providing learners with time to engage in careful on-line planning may be an important factor in increasing accuracy of production.

*Planning instructions.* Strategic planning research has also looked at the impact of different strategic planning instructions on learners’ subsequent performance. Some studies (e.g., Foster & Skehan, 1996; Skehan & Foster, 2005) have evaluated whether providing detailed suggestions regarding the use of strategic planning time as part of the planning instructions would positively impact learners’ production. Other studies (e.g., Foster & Skehan, 1999; Mochizuki & Ortega, 2008; Sangarun, 2005) have examined whether content-focused or grammar-focused instructions would change learners’ production. In the studies by Foster and Skehan (1996) investigating the level of detail provided in instructions, the detailed versions included suggestions such as, “Think of ways to make sure your listener won’t get lost”, and “Think what grammar you need to do the task”. Regarding linguistic complexity, both studies reported that the planning done by learners who had been given detailed instructions was more beneficial than the planning for which no detailed suggestions had been provided. The results for accuracy were rather contradictory, in that the most accurate production was found for the
undetailed group in the 1996 study, but for the detailed group in the 2005 study. Skehan and Foster (2005) did not provide any explanations for this discrepancy. However, it may be due to the differences between the two studies in the type of tasks used and proficiency levels of the learners. In the 1996 study three tasks were used, a personal information exchange task, a narrative task, and a decision-making task; whereas the 2005 study only used decision-making tasks. In addition, learners in the 1996 study were pre-intermediate; in the 2005 study, learners were intermediate. These factors must, of course, to be examined in the future studies.

As for the content versus grammar focus, studies have revealed rather mixed results. Mochizuki and Ortega (2008) found that learners with form-focused strategic planning, which provided grammar explanations of the form that needed to be used for the task performance, showed significantly more accuracy and complexity on the instructed forms compared to those with no planning (no content-focused group was used in this study). This finding apparently provides support for grammar focus planning, but it may also indicate that the results may not have been related to planning per se, but rather to the fact that the learners in the grammar-focused planning group were given explanations. Foster and Skehan (1999) and Sangarun (2005) did not find any significant differences in the complexity or accuracy of learners’ L2 production when participants were given content- versus grammar-oriented instructions. However, Sangarun did find that instructions for strategic planning which involved both meaning and form were beneficial: When learners were given information on 1) how to plan meaning, 2) useful forms for the upcoming task performance along with targetlike examples, and 3) how to
combine content and forms, they consistently outperformed participants in a no planning condition.

Sangarun’s (2005) study, which also gathered qualitative data, sheds light on some of the results in the previous studies. By using think-aloud protocols, she was able to collect data regarding how learners used their strategic planning time and discovered that across all conditions (meaning-focused, form-focused, meaning-plus form-focused), the participants showed a clear tendency to attend to content rather than form. This concurs with the argument made by VanPatten (1996), which stated that learners may tend to focus on meaning rather than form and by Schmidt (1990), that indicated learners may focus, either deliberately or non-deliberately, on items the task instructions do not specify. Many researchers have stressed the need for more studies of what learners do during planning (e.g., Batstone, 2005; Ellis, 2009; Mochizuki & Ortega, 2008; Ortega, 1999, 2005; Sangarun, 2005; Skehan & Foster, 2005). However, thus far, few (Ortega, 1995, 2005; Sangarun, 2005) have explored this area.

Learners’ perceptions. Just as Sangarun’s (2005) study provided insights into what learners were actually doing while engaged in strategic planning, investigations of learners’ perceptions of task difficulty may also need to be examined to understand the impact of strategic planning. The aforementioned attentional models, the single-resource model (Skehan, 1998; Skehan & Foster, 2001) and the multiple-resource model (e.g., Robinson, 2007; Wickens, 1989, 1992), predict that strategic planning may reduce the attentional demands involved in transacting a task and facilitate focus on form. Accordingly, it seems reasonable to expect that when planning time is given, learners
might perceive the task at hand to be easier compared to instances where they are not given time to plan. Considering that learners’ perceptions may be affected by their intelligence, aptitudes, and motivations, Robinson (2001) cautions against equating learners’ perceptions of task difficulty with the actual cognitive demands of a task; however, there is some evidence for an association: Robinson (2001) and Baralt (2010), for example, found that learners rated a cognitively more complex task as being more difficult.

At the same time, though, several studies investigating the effects of strategic planning (Elder & Iwashita, 2005; Tavakoli & Skehan, 2005; Wigglesworth, 2001) have reported mixed results regarding learners’ perceptions of task difficulty. Tavakoli and Skehan (2005) found that planners perceived tasks to be significantly easier than non-planners did. Elder and Iwashita (2005) showed no significant differences in learners’ perceptions of difficulty in terms of the availability of strategic planning time. In Wigglesworth (2001), learners who engaged in familiar or structured tasks perceived them to be easier with strategic planning than without planning. On the other hand, when engaged in unfamiliar or unstructured tasks, they reported the tasks were more difficult with planning. Ortega (1999, 2005), who interviewed learners about strategic planning, revealed that some learners experienced added pressure due to the planning. Their comments included, “a little bit more tense to do it, because getting to write notes [during planning] ‘oh I have to remember more things’”, and “When you prepare you get worried. You get nervous that you are going to mess up and then you probably do. That’s what I think” (Ortega, 2005, p. 90). Taken together, such findings pose some question to the
underlying theoretical claim of task planning studies. These studies suggest that, under some conditions, strategic planning may not function as theoretically assumed and may even act quite contrary to prediction. It appears that there is a need for future studies in order to address this issue.

*Summary of Task Planning and Task Performance*

Previous studies on strategic planning revealed that several factors should be considered when investigating the effects of strategic planning on bringing learners’ attention to form-oriented aspects of production during task performance. While these studies provided important insights, in order to fully understand the effects of strategic planning, researchers have called for more introspective studies (e.g., Batstone, 2005; Ellis, 2009; Mochizuki & Ortega, 2008; Ortega, 1999, 2005; Sangarun, 2005; Skehan & Foster, 2005). Further introspective studies will bring better understanding of what learners actually do during strategic planning and how learners perceived the impact of strategic planning on their task performance. The results of these studies, therefore, will significantly contribute to strategic planning research.

*Strategic Planning, Noticing, and L2 Development*

*Issues with Previous Strategic Planning Studies on Task Performance*

One of the major limitations of previous task planning studies on task performance, Ellis (2005, 2009) has argued, is that they do not provide enough evidence related to L2 development. Ellis (2009) states:

In order to theorize the relationship between planning and the acquisition of linguistic resources, it is important to distinguish three senses of acquisition:
(i) the acquisition of new linguistic features, (ii) the restructuring of existing L2 knowledge, (iii) the development of greater control (accuracy) over existing linguistic features… the effects of planning are likely to be seen in restructuring and greater control (p. 504).

Skehan (1998) has argued that task planning can assist restructuring by the effect it has on the linguistic complexity of learners’ L2 production. In other words, he assumed that planning leads to more complex language, and then more complex production will lead to acquisition. This argument, however, is called into question by Batstone (2005):

If planning is to lead to payoffs in terms of learning, learners need (amongst other things) to push their output by using challenging, relatively complex language… But most of the extant research on planning does not convincingly demonstrate that planning necessarily leads to output pushing. To date, scholars have relied almost entirely on the presentation of abstracted quantitative data as evidence for the impact of planning on learners’ language production. But this kind of data fails fully to demonstrate that learners are in fact ‘pushing’ their language to the extent that they could or (perhaps) that they should be doing in order to maximize their learning (p. 279).

Simply put, as Batstone argued, research on planning has not persuasively shown that planning leads to maximal output pushing, and the link between planning and L2 acquisition via output pushing must yet to be proved. Moreover, for accuracy, because the majority of the planning studies does not collect baseline data and have employed global measures instead of measures focused on specific linguistic features, it is difficult
to observe the effects of planning on the development of greater control or accuracy (Ellis, 2005). Based on these arguments regarding the scarcity of evidence, Ellis (2005) suggested that it may be necessary for research in this area to incorporate standard experimental designs involving a pre-test, a treatment, and posttests.

In addition to these concerns with demonstrating L2 development, several researchers (e.g., Ellis, 2005; Mackey, 2012; Robinson, 2001; Robinson & Gilabert, 2007) have also expressed that future task planning studies must extend the scope of their focus to include the actual processes learners go through in incorporating input, more specifically the noticing and acquisition of new linguistic information. Production-based studies have previously focused on whether planning helps learners in accessing “their own implicit and explicit knowledge of the L2 for use in production” (Ellis, 2005, p. 7) but not on input incorporation.

**Implications from the Two Theoretical Models of Attention**

Skehan (1998; Skehan & Foster, 2001), in his single-resource model, has not explicitly addressed the connection between pre-task planning and acquisition of new linguistic information; however, some inference between them can be made. Single-resource model maintains that attentional resources are limited; therefore, when a significant amount of attention is required to simply transact a task, there would be an insufficient amount of attention available to allocate to its language. Base on this argument, it can be logically assumed that when learners are given time for strategic planning, which Skehan argued reduces the attentional demands from completing a task, the chance that learners will notice and develop new information will increase.
Theoretical claims from Robinson and Gilabert (2007) and Robinson (2010), which are built on multiple-resource model, indicated more than one perspective relating to the connection between task planning and noticing/L2 development.

In Robinson and Gilabert, referring to Robinson’s (e.g., 2001, 2003, 2005) Cognition Hypothesis, they state:

It predicts that along resource-directing dimensions, and in general too along resource-dispersing dimensions, that more interactive complex tasks will result in greater amounts of interaction, and negotiation for meaning. It also claims, following Long (1996), that such negotiation provides a context for attending to problematic forms in the input and output, and additionally that on complex versions of tasks there will be greater attention to, and uptake of forms made salient during provision of reactive Focus on Form techniques such as recasts (see Doughty 2001; Long 2007; Long and Robinson 1998) (Robinson & Gilabert, 2007, p. 167).

In this statement, they did not make a distinction between resource-dispersing and resource-directing dimensions regarding the prediction for noticing of interactional features. Therefore, when the complexity of tasks is increased in resource-dispersing dimensions (such as learners not being given time to plan prior to interactive task performance), it is predicted that they are likely to seek information from input and have more negotiations, resulting in greater attention to and uptake of recasts. In a more recent statement, however, Robinson (2010) changed his prediction related to resource-dispersing dimension, and has argued that making a task more complex in resource-
dispersing dimensions, such as removing planning time, “promotes not noticing of language code and interlanguage development of new linguistic, conceptual form-function mappings, but rather consolidation and fast real-time access to existing interlanguage resources” (p. 252). This latter statement, therefore, does not support his earlier hypothesis that noticing would occur more among learners without planning. In addition, it does not articulate that by decreasing the cognitive demands of tasks in resource-dispersing dimensions, noticing would increase. However, his writings regarding the manipulation of resource-dispersing dimensions’ affecting learners’ task-performance, specifically on complexity and accuracy, provides some support to the argument that decreasing task demand in resource-dispersing dimensions can encourage learners’ noticing and therefore L2 acquisition.

Robinson (e.g., 2001, 2003, 2005, 2007) argued that when task demand is increased in resource-dispersing factors (e.g., withdrawing planning time), it negatively affect all aspects (fluency, accuracy, and complexity) of L2 performance. This argument can be taken to assume that the provision of strategic planning may help avoid negative consequences to learners’ attention to language during task performance; therefore, as a result, it supports learners’ noticing and acquisition.

In sum, implications from the two models of attention predict that provision of strategic planning would support learners’ noticing and thus L2 acquisition (see Figure 1).
Thus far, only one empirical study, Romanova (2010), has addressed the link between strategic planning and noticing/L2 development. All of the learners in the study were given recasts on the target form, the Russian third-person singular form, during a picture description activity; but they were divided into three groups: strategic pre-task planning, within-task planning, and no planning. The learners with strategic planning
were given five minutes of planning time prior to the picture description task, but during the task they were given a limited amount of time to describe each picture; therefore, they were under time pressure. The learners assigned to the within-task planning group were not given time to plan prior to task performance; however, they were allowed to perform the description task at their own pace (i.e., without time pressure). The learners with no planning time were not given planning time prior to task performance and they had to perform the task under time pressure. Upon the completion of the recast treatment, the learners participated in an exit questionnaire where they were asked five questions, and then they completed immediate posttests composed of two oral tasks, a picture description task and an answer-the-question task. Two weeks later, the delayed posttest, in the same format as the immediate posttest, was administered to the learners.

Descriptive statistics from both the immediate and delayed posttests suggested that 1) learners with within-task planning demonstrated the most accurate performance , 2) learners with strategic planning were more accurate than learners with no planning, and 3) learners with within-task planning were more accurate than learners with no planning. A statistically significant difference was found only between the latter pair, within-task planning versus no planning, for both the immediate and delayed posttests. As for learners’ noticing, Romanova reported that, “noticing the form during the task performance was more prominent both the PP [strategic planning] and OP [within-task planning] groups (90% and 80%) than for the NP [no planning] group (38%)” (p. 862). To summarize the result regarding strategic planning, even though Romanova’s posttest results did not show a significant difference, the results did lean in the direction that the
learners with strategic planning were more accurate than the learners with no planning; therefore, the results on posttests and noticing regarding strategic planning, in general, support the implications drawn from the two models of attentions regarding strategic planning.

It is important to mention that there were some methodological limitations in Romanova’s study. First, the noticing measure she used was a questionnaire, a type of offline measure which suffers validity concerns. Typically, in interaction studies, because it is not possible to implement online noticing measures, researchers use an offline measure called stimulated recall to gauge learners’ noticing. As discussed earlier in this chapter, this technique uses video- or audio- recordings as stimuli to serve as memory aids for the recall; therefore, it reduces the memory decay concerns inherent in offline measures, and it allows researchers to obtain more accurate and complete responses compared to offline measures, such as questionnaires. Romanova’s questionnaire did not provide support for memory retrieval, and no additional noticing measures were used; therefore, her “noticing” data may be inaccurate and incomplete. Second, in her study, “noticing” was not operationalized clearly. What I was able to gather from her article was that she used only one of the five questions in the questionnaire, specifically “What did you notice during your interaction with the researcher?” (p. 854), to collect data on noticing. Without a clear and well defined operationalization of noticing, it is impossible to evaluate the credibility of the results.
Considering the limitations of Romanova’s study and that it is so far the only study to investigate the relationship between task planning and noticing/L2 development through recasts, this topic clearly warrants additional research.

Summary on Strategic Planning, Noticing, and L2 Development

The implications drawn from the two models of attention predict that the provision of strategic planning supports learner’s noticing and L2 acquisition. Only one study has so far empirically tested the connection between strategic planning on the one hand and noticing and L2 development on the other. Romanova (2010) examined whether the availability of planning time would alter the way learners benefited from recasts in terms of their noticing and L2 development. Her results showed that learners with strategic planning reported significantly more noticing of the target form compared to learners without planning. Also, the descriptive statistics from both immediate and delayed posttests suggested that learners provided with strategic planning time were more accurate than learners with no planning (however, it did not reach significance). These results, therefore generally confirm the implications from the existing theoretical models. However, of course, more research is necessary to further verify the connection, and it is important that these future studies use sound methodologies to evaluate this issue.

Research Questions

This Dissertation addressed the following questions:

Research Question 1

Do learners who receive recasts following strategic planning show greater L2 development than learners who have no planning time available?
Research Question 2

Do learners who receive recasts following strategic planning show greater noticing of L2 linguistic forms and pragmatic factors than learners who have no planning time available?

In this current study, L2 development was operationalized as improved accuracy in oral, written production, and written recognition tests in relation to the linguistic target (see the Method chapter). The operationalization of noticing of L2 linguistic forms and pragmatic factors are also provided in the Method chapter.

Focusing on these questions, this dissertation was designed to reflect and expand on what had been argued or investigated in the areas reviewed in this chapter, 1) Noticing, 2) Recasts, and 3) Strategic Planning, Noticing, and L2 Development. Regarding noticing, Schmidt has argued that noticing of different areas of language may necessitate different degrees of attention. Differing from some morphological rules (which may be rather transparent to learners), learning pragmatics often involves, “new interpretative assessment of the world” (Schmidt, 1993b, p.34); therefore, noticing of new pragmatic systems may require a deeper degree of focal attention, or deliberate focal attention. In this dissertation, I have selected the pragmatic rules associated with Japanese verbs as the target form of the study. The results of this dissertation, therefore, hope to provide some insights into his argument.

Previous studies in recasts have shown that several factors may affect the way learners notice and benefit they may receive from recasts. Recently a few studies have begun examining different task variables as another factor. One of these studies
(Romanova, 2010) examined the task variable, strategic planning, which has been widely examined in task-bases learning research. In this dissertation, the variable was further examined in relation to noticing and L2 development; therefore, this current study’s results will provide additional empirical data for how the task variable influences the efficacy of recasts. The results regarding noticing and L2 development will also account for what Ellis (2005, 2009) called one of the major limitations of previous task planning studies. Ellis has argued that future task planning studies must provide evidence related to L2 development. This present study will directly address the topic.

Task planning researchers have also called for more introspective studies in order to better understand the process in which learners engage when they are provided with strategic planning time. Thus far, only a few previous studies have investigated learners’ perception, and further investigation is warranted. In this dissertation, several introspective measures are used to address the second research question, learners’ noticing. Learners’ comments collected through these measures are also utilized to explain the quantitative results used to answer the two research question. These comments will shed light on the issue of learners thought processes and perceptions toward the use of planning.
Chapter 2: Method

Participants

Fifty native or advanced speakers of English who were learning Japanese in second- and third-semester university-level Japanese classes participated in this study. Twenty-seven participants were women and the remaining 23 were men. Their ages ranged from 18 to 23, averaging 19.59. Their reported native language(s) were English (n = 30), Korean (n = 10), Chinese (n = 9), German (n = 1), Portuguese (n = 1), and Spanish (n = 1). Twenty-two participants were recruited from second-semester intensive Japanese classes at Georgetown University (GU) and 28 participants were recruited from third-semester non-intensive Japanese classes at The George Washington University (GWU). The GU students were receiving six hours of Japanese instruction per week, whereas the GWU students were receiving four hours of instruction per week. At the time of data collection, the approximate total number of hours of Japanese instruction the participants had received at their respective universities was 126 hours for each GU student and 140 hours for each GWU student.

In order to recruit learners of Japanese from these universities, first, I contacted the Japanese program coordinator at each university and received approval to visit the Japanese classes. During my visits to the classes, I informed the students that I was interested in observing how Japanese was learned using one-on-one communicative activities with a native speaker and that they would be asked to meet twice with me, the researcher, individually. The more specific purposes of the research were not revealed to the participants. I then handed out sheets of paper to the students and asked them to write
their name, contact information, and the dates/times of their availability if they were interested in participating in my study. I collected the papers with their information and scheduled one-on-one data collection sessions via email. All of the students \( n = 62 \) who attended our first meeting signed consent forms and completed pretests. Based on the results of the pretests, 11 students were eliminated using criteria to be discussed below. I explained the purpose of the study and the target form to those 11 students, and they did not participate in the rest of the experiment. The remaining 51 proceeded to the second meeting, during which they received treatments and completed posttests. Later, during the data analysis phase, it was found that one of the participants’ video data files did not contain sound; therefore, only the remaining 50 participants’ data were used for analysis.

*Linguistic Target*

The linguistic target of this study was the pragmatic rules associated with Japanese *give* verbs. In Japanese, there are two verbs which correspond to the English verb *give*, namely, *ageru* and *kureru*. As explained by Tsujimura (2007), *ageru* is used when the speaker or anyone in speaker’s in-group is the agent, or giver. In other words, *ageru* is used when the speaker herself/himself or a close relation (e.g., family member, friend) gives something to someone (see Example 2).

**Example 2. Examples of sentences with *ageru***

\[
\begin{align*}
\text{Watashi-wa} & \quad \text{sensei-ni} & \quad \text{purezento-o} & \quad \text{ageru}. \\
\text{Self-TOP} & \quad \text{teacher-DAT} & \quad \text{present-ACC} & \quad \text{give} \\
\text{I give (my) teacher a present.}
\end{align*}
\]

\[
\begin{align*}
\text{Haha-wa} & \quad \text{sensei-ni} & \quad \text{purezento-o} & \quad \text{ageru}. \\
\text{Mother-TOP} & \quad \text{teacher-DAT} & \quad \text{present-ACC} & \quad \text{give} \\
\text{(My) mother gives (my) teacher a present.}
\end{align*}
\]
On the other hand, *kureru* is used when the giver is someone other than the speaker or the speaker’s in-group. That is, *kureru* is used in situations where someone with whom the speaker is not closely related (i.e., person in out-group such as the speaker’s teacher, boss, guest, etc.) gives the speaker, or the speaker’s in-group, something (see Example 3).

Example 3. Examples of sentences with *kureru*

<table>
<thead>
<tr>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensei-wa  watashi-ni  purezento-o  kureru.</td>
</tr>
<tr>
<td>Teacher-TOP  Self-DAT  present-ACC  give</td>
</tr>
<tr>
<td>(My) teacher gives me a present.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensei-wa  haha-ni  purezento-o  kureru.</td>
</tr>
<tr>
<td>Teacher-TOP  mother-DAT  present-ACC  give</td>
</tr>
<tr>
<td>(My) teacher gives (my) mother a present.</td>
</tr>
</tbody>
</table>

This in-group and out-group relationship is not static and changes according to the situation. This relationship can be conceptualized as a series of overlapping circles (or they are on a continuum of in-groupness), and directionality of a giving action in relation to the in-group and out-group relationship determines which verbs (*ageru* or *kureru*) should be used (see Figure 2). For example, when the speaker (or ‘self’) is the agent and her family member is the receiver (i.e., I give my sister); the speaker becomes the in-group and the family member becomes the out-group; therefore, *ageru* is used. When her family member is the agent and her friend is the receiver (i.e., My sister gives my friend), the family member becomes the in-group and the friend becomes the out-group; therefore, again *ageru* is used. Kureru can be described in the opposite manner. For example, when her family member is the agent and the speaker is the receiver (i.e., My sister gives me), *kureru* is used. When her friend is the agent and her family member is the receiver (i.e., My friend gives my sister), again, *kureru* is used.
Summarizing reports from previous studies on the first language acquisition of Japanese, Clancy (1985) has noted that in children’s language, *ageru* emerges first and then *kureru*. Horiguchi (1979), in a detailed case study, showed that a child two years of age had difficulties correctly using *ageru* and *kureru*, but that the pattern of errors was quite asymmetrical. *Ageru* was rarely used incorrectly, whereas *kureru* was incorrect almost half the number of times it occurred. Based on these L1 data, Horiguchi proposed that *ageru* is easier to acquire for children, especially in the case when the agent is the speaker, perhaps because with *ageru* the roles of speaker, topic, and agent are combined.

At Georgetown University and The George Washington University, *ageru* is taught in beginning-level classes along with the Japanese dative construction. *Kureru* is taught at a later stage, in upper-beginning or intermediate classes. The pragmatic function
of *give* verbs is not mentioned when *ageru* is introduced. When *kureru* is covered in class, the use of these two verbs with respect to the in-group/out-group concept is explicitly explained to students. Thus, it is common to observe beginning learners, who are taught only *ageru*, creating erroneous sentences as in the example below⁴.

Example 4. An example of an erroneous *give* verb sentence

* Sensei-wa watashi-ni purezento-o *ageru.*
Teacher-TOP self-DAT present-ACC give
(My) teacher gives me a present.

In this example, since the giver is an out-group person and the receiver is the speaker (in-group person) the correct verb choice is *kureru*; however, the verb *ageru* is used instead. It violates the pragmatic rules of the *give* verb. Even after the explanation of pragmatic rules is given to learners, it is still common to see the type of error exemplified above among Japanese language learners.

Another common mistake related to Japanese *give* verbs is the use of the word *moraru*, which means ‘to receive’. *Morau* is also bound by the in-group and out-group concept, and it can be only used for the situation where a person in the in-group is the receiver and a person in the out-group is the giver. Therefore, *morau* cannot be used when the receiver is a person in the out-group and the giver is a person in the in-group (see Example 5).

---

⁴ I observed four beginning Japanese classes. When students made this error, the teacher usually ignored it. However, one time, a teacher recasted the error. No explanation for the correction was provided to the student.
Example 5. Examples of a correct and an erroneous receive verb sentence

Watashi-wa sensei-ni purezento-o morau.
Self-TOP teacher-DAT present-ACC receive
I receive from (my) teacher a present.

* Sensei-wa watashi-ni purezento-o morau.
Teacher-TOP self-DAT present-ACC receive
(My) teacher receives from me a present.

In addition, morau and kureru, can describe the same giving and receiving situation (see Example 6).

Example 6. Examples of morau and kureru sentences

Watashi-wa sensei-ni purerzento-o morau.
Self-NOM teacher-DAT present-ACC receive
I receive from (my teacher) a present.

Sensei-wa watashi-ni purezento-o kureu
Teacher-NOM self-DAT present-ACC give
My teacher gives me a present.

Scope of the Linguistic Target Addressed in the Experiment

In order to facilitate the participants’ learning of the relationship between the give verbs and in-group/out-group distinction during the treatment session of the current experiment, I decided to focus on the usage of the verbs only in the following limited situations: 1) Self (in-group) and acquaintance (out-group), and 2) family members (in-group) and acquaintance (out-group). On the posttests, however, in addition to those situations, friends (in-group) and celebrities (out-group) were included in order to test participants’ abilities to generalize what they had learned through the treatment.
Participant Criteria

For this present study, I only included learners who 1) knew and could use ageru, 2) did not know kureru, and 3) did not know about the give verbs’ in-group/out-group distinction. Also, it is crucial to mention that this in-group/out-group distinction does not exist in any of the present study’s participants’ first languages (Chinese, English, German, Korean, Portuguese, Spanish) or in any other second languages they had previously studied (Chinese, English, German, French, Korean, Latin, Spanish, Farsi). Therefore, the chance that their previous language experiences might influence their learning of these pragmatic rules was minimized.

Procedure

The experiment was conducted over two days for each participant. On the first day or at the first session, the participant completed 1) a consent form and 2) pretests. On the second day or at the second session, the participant completed 1) a treatment, 2) a stimulated recall interview or posttests, and 3) an exit questionnaire. All activities were conducted in one-on-one settings with the researcher (or with a research assistant, in the case of the stimulated recalls). The time lapse between the first and second days was between one and two weeks. A visual representation of the procedure is found in Figure 3.
77 participants completed the two-day experiment; however, one of the 15 participants’ data were not usable for data analysis; therefore, the number of participants whose data were included in the analysis was 14 for this group.

Figure 3. Experiment design.
Setting

At both Georgetown University and The George Washington University, much of data collection took place using a 19-inch laptop. The laptop was equipped with audio and video recording software, Record Pad and Debut Video Capture. For audio recording, a Logitech USB microphone was attached to the laptop; and video recording was conducted using a web cam and a Logitech USB microphone attached to the laptop. When the laptop was used to show the participants the video recording of the treatment task interactions during the stimulated recall interview, the video files were played through Windows Media Player installed on the laptop. The laptop was also used to display the PowerPoint slides used for 1) oral tests instructions and 2) written posttests (other materials used for the experiment will be described in the section called Materials in this chapter).

At Georgetown University, a small linguistic lab, created for data collection of linguistics research, was used for all the experiment meetings. The lab contained 1) a few desktop computers on a long counter attached on one side of the wall and 2) two chairs. The laptop was placed in the open space of the counter and the two chairs were placed in front of the laptop during the experiment.

At The George Washington University, a large computer lab was used. The lab was located in the basement of the library. It was open to the university students during the data collection; however, very few students entered the lab, and distraction from other students was quite minimal. The laptop used for the experiment was placed at the far corner of the lab to increase the sense of privacy for the participants.
Day One

The researcher handed out the consent form (see Appendix A) and directed the participant to read it and to ask the researcher if he/she had any question. The participant then signed the form if he/she agreed to participate in the study.

Pretests. After the consent form was signed, the participant completed three pretests in the order presented here: 1) an oral task, 2) a written fill-in-the-blank production task, and 3) a written fill-in-the-blank recognition task. The recognition task was placed at the end because the task provided target words in a word bank. It was feared that if the recognition task occurred before the other two tasks, the words in the word bank might prime the participants’ responses on the production tasks (e.g., by alerting them to the existence of kureru).

As part of the instructions for the oral pretest, the learner completed two trial items in the same format as the items on the oral pretest. The trial items were presented using PowerPoint slides with an explanation of how the task should be completed (see Appendix B). Along with the trial items, the people appearing on the oral test were introduced using the same pictures used in the test (see Appendix C). It was explained that the person referred to by watashi (i.e., I, me, my) and the family members on the test represented the participant and the participant’s family, respectively. During these instructions, any questions regarding the oral task were resolved. Upon the completion of the trial items, the participant received materials of the oral pretest. During this test, no feedback regarding the correctness of the participant’s answers was provided. The participant’s answers were audio recorded on a computer.
Following the oral pretest, the participant was presented with an answer sheet and PowerPoint slides containing the questions for the written production test. On the first slide, there was a short set of instructions for the test which warned the participant not to return to previous slides (see Appendix D). The participant recorded his/her answers on the answer sheet in pen or pencil. Upon the completion of the written production test, the written recognition test was conducted. The participant was given another answer sheet (this time with a word bank) and PowerPoint slides with the test items. Again, the first slide with the instructions warned the participant not to return to previous slides. The participant recorded his/her answers on the answer sheet in pen or pencil. During this written test phase, the researcher sat next to the participant to ensure that everything was completed as instructed.

The pretests had two functions: 1) to ensure that the participants had knowledge of *ageru* and could use *ageru* and 2) to ensure that the participants did not have knowledge of *kureru*, another *give* verb. For each of the three pretests, six items were designed to elicit *give* verbs. Participants who failed to supply *ageru*⁵ on more than one of these six items in any of the pretests, or participants who used *kureru* at all on any of these six items, were excluded from the present study. In total, 11 learners were excluded on the basis of their pretest performance. The remaining 51 participants were randomly assigned among four treatment groups. The random assignment was conducted independently at each university in order to ensure that the same proportion of

---

⁵ There were three learners who used *akeru* instead of *ageru* on *ageru* items. In Japanese, *ke* and *ge* are quite similar in writing, け and げ respectively, and some learners have difficulty hearing the difference between the two. Therefore, I did not count *akeru* to be incorrect.
participants from each school would be placed in each experimental group. This helped to control for the possible influence of previous instruction the participants had received at each university.

**Day Two**

*Treatment.* The participants were randomly assigned among the four treatment groups: (1) the strategic planning and posttest group (+P/T), (2) the no strategic planning and posttest group (-P/T), (3) the strategic planning and stimulated recall group (+P/SR), and (4) the no strategic planning and stimulated recall group (-P/SR). The number of participants in each group follows: 15 for the strategic planning and posttest group (+P/T), 16 for the no strategic planning and posttest group (-P/T), 10 for the strategic planning and stimulated recall group (+P/SR), and 10 participants for the no strategic planning and stimulated recall group (-P/SR). The discrepancy among the numbers of participants in each group was mainly due to the small number of participants who remained after the pretests. Ideally, an equal number of participants in each group is desired; however, more numbers of participants were randomly but purposefully assigned to the posttest groups (+P/T, -P/T) than the stimulated recall groups (+P/SR, -P/SR). This occurred because larger numbers of participants were necessary to have accurate results from statistical analyses of their test scores between the groups. As for the stimulated recall groups (+P/SR, -P/SR), larger numbers of participants were also necessary for accurate statistical analysis; however, on the chance that no statistical significance was reached due to the small number of participants, I expected that due to the qualitative nature of the data gathered, the results would still provide useful insights.
Participants in each group completed two treatment tasks. In next few paragraphs, the specific procedures used during the treatment for the planning groups (+P/T and +P/SR) will be described, followed by a description of the specific procedures used for the no-planning groups (-P/T and -P/SR)\(^6\).

Prior to the first treatment task, the participants in the planning groups (+P/T and +P/SR) received verbal instructions for thinking aloud and engaged in a practice task to familiarize themselves with the process of thinking aloud (Appendix E). The practice task involved thinking aloud while solving math problems. The participants were instructed to simply state whatever came to mind in either Japanese or English.

Upon the completion of the think-aloud practice, the participants proceeded to the first treatment task, Task One. First, each participant completed instructions and trial tasks for treatment task (see Appendix B), then received written planning instructions. The participant was asked to read the instructions aloud. He/she also received task materials for Task One. The participant was told to take 10 minutes to plan for the upcoming task. The participant was also instructed to think aloud during the planning activity, just as he/she had done with the math problems. The researcher left the room and the participant was allowed to complete the task in private. The think-aloud was audio recorded. The participant was told that the blank space below the planning instructions could be used as a note-taking sheet if that would help with the planning, but was asked not to write anything on the task materials and was warned that the note-taking sheet would not be available during the upcoming task performance.

\(^6\) Here, P stands for planning, T stands for test, and SR stands for stimulated recall
After the planning, the participant completed Task One with the researcher. Whenever a participant produced an error with a *give* verb, the researcher provided a partial recast with declarative intonation. The type of recast was kept constant because previous research has shown that various features of recasts are differentially associated with noticing and L2 development. In order to keep the participant’s focus primarily on meaning in the context of a communicative activity, each partial recast included not only the *give* verb but also the object of the verb. For similar reasons, a partial recast was also provided whenever a participant incorrectly read the items on the task materials.

Following Task One, the participant engaged in another planning session before proceeding to Task Two. The procedures for Task Two were the same as for Task One. Each participant’s interactions during Task One and Task Two were video recorded.

The participants in the no-planning groups (−P/T and −P/SR) did not engage in the planning activities. Therefore, when these participants met with the researcher, they completed instructions and trial tasks for treatment task (see Appendix B) and were simply instructed to engage in a speaking activity similar to what they had done during the first meeting and were given brief instructions that directed them to focus on 1) making appropriate choices of gifts and money amounts, 2) communicating naturally and fluently, and 3) using language that was grammatically correct (see Appendix F). The instructions were essentially the same as those given to the participants in the planning groups during the planning activity. The participants in the no-planning groups read the instructions aloud and then received the task materials for Task One. Just as in the planning groups, whenever these participants produced errors on *give* verbs while
carrying out the task, the researcher provided partial recasts with declarative intonation. Partial recasts were also given whenever the participants incorrectly read the items in the task materials. Task Two was completed in the same manner. Each participant’s interactions during Task One and Task Two were video recorded.

Posttests. Immediately following their treatment sessions, the participants in the posttest groups (+P/T and -P/T) completed the three posttests in the order presented here: 1) the oral task, 2) the written fill-in-the-blank production task, and 3) the written fill-in-the-blank recognition task. The procedure of the posttests was the same as that of the pretests. The participants’ answers in the oral test were audio recorded. Their answers on the written tests were recorded in pen or pencil on the answer sheets.

Stimulated recall interview. Immediately following their treatment sessions, the participants in the stimulated recall groups (+P/SR and -P/SR) met with a research assistant and participated in individual stimulated recall interviews. Prior to each participant’s interview, the research assistant provided a set of written instructions (adapted from Egi, 2004) and asked the participant to read them aloud. The instructions stated that 1) the participant was going to watch the task interactions that he/she had just had, 2) the researcher (i.e., the one conducted the treatment) wanted to know what the participant had been thinking at the time he/she was doing the tasks, 3) when the research assistant paused the video, the participant would need to articulate in English or Japanese whatever he/she had been thinking at that time in the interaction, 4) if the participant did not remember what he/she had been thinking at that time or if he/she had not been
thinking anything in particular, he/she should say “I don’t remember” or “I wasn’t thinking anything” (see Appendix G).

The research assistant showed the participants their respective video recordings from Task One and Task Two, stopping the recording at the end of each item in the two tasks. Since each task had 16 items, each participant was asked about his/her thoughts 16 times for each task, for a total of 32 times. During the interview, the participants were continually reminded to report only thoughts that had occurred at the time of the task interaction. Their stimulated recall comments were audio recorded.

Exit questionnaire and biodata. After either the posttests or the stimulated recall interview, each participant was given an exit questionnaire which included 1) six questions designed to assess noticing of the pragmatic rules associated with give verbs, 2) a question regarding the perceived difficulty of the treatment tasks, and 3) a question on the perceived helpfulness of planning; and he or she answered the questions in pen or pencil. Upon completion of the exit questionnaire, the researcher or the research assistant interviewed the participants regarding the following background variables: 1) gender, 2) age, 3) native language(s), 4) whether they had visited Japan and, if so, when and for how long, 4) whether they had lived in Japan and, if so, when and for how long, and 5) foreign language experiences, including which language(s) they had studied, for how long, where, and to what proficiency level (see Appendix H for the biodata questionnaire). Their answers were written down by the researcher or the research assistant on the biodata questionnaire.
Materials

Treatment Session

This section describes two kinds of materials used in the treatment session: 1) the treatment tasks and 2) the strategic planning instructions. For both, the materials used for piloting are explained first, followed by a description of the actual materials used during the study.

Piloting of treatment tasks. For the piloting of the treatment tasks, three interaction tasks designed to elicit the give verbs were created and trialed with seven students at the same course level as the participants for the actual study. All of them were decision-making tasks based on Japanese money- and gift-giving customs. For Task One, the participant was first asked to read the situation below:

You are a 28-year-old Japanese actor who has been working in the US. Two weeks ago, you were in a car accident and were hospitalized for a week with several broken bones. During your stay in the hospital, your friends came to visit you. Some of these visitors gave you a gift or money (Ominai), as is the custom in Japan. Now you are out of the hospital, and you need to send Kaiki Iwai, celebrating your recovery, to the people who gave you Ominai. You remember who came to the hospital and have been able to come up with the list of people. However, you don’t remember who gave you Ominai. One by one, using your lists, ask your mother (your Japanese partner) who gave you Ominai and choose an appropriate Kaiki Iwai gift for each person. Tell your mother your reasons for each choice.
After reading the situation above, each participant was given a list of 12 celebrities with corresponding pictures and pictures of 15 gifts (see Example 7). The participant was reminded that these celebrities were his/her friends. The task was to engage in interaction in Japanese, producing sentences which contained the give verbs and providing an explanation of each selection. Performing this task with linguistic accuracy required creating sentences using both *ageru* and *kureru*.

**Example 7. Examples of items in the pilot treatment task**

**Visitors:**

_Gave you *Omimai*?_

- Paris  
  - Y / N  
  - Gift________________

- Tom  
  - Y / N  
  - Gift________________

For Task Two, the task format was the same as Task One; however, the situation given for this task included people to whom the participants would need to show respect (e.g., President Obama, French President Sarkozy; cf. friends in Task One). Therefore, it
was designed to elicit 1) *sashiageru*, a humble expression of *ageru*, and 2) *kudasaru*, a respectful expression of *kureru*. *Ageru* and *kureru* were considered errors in this task. Task Three included both friends and people to whom participants would be expected to show respect. For this task, therefore, it was necessary to use all of the *give* verbs appearing in Task One and Task Two, *ageru*, *kureru*, *sashiageru*, and *kudasaru*. All the pilot participants engaged in the tasks in the same order and completed all of the tasks in one meeting.

Observations and comments from the piloting participants revealed several problems with the piloted treatment tasks. First, some of them mentioned that the tasks felt artificial and that reading the situation and performing the task were rather confusing. Second, the time required to complete the three tasks was too long considering that participants in the actual data collection would be asked to engage in additional posttests or stimulated recall protocols, not to mention that some participants would also be given time to plan prior to performing the tasks within the same meeting. Third, having exposure to four words (*ageru*, *kureru*, *sashiageru*, and *kudasaru*) in one meeting seemed overly difficult and confusing. Fourth, the task design required the piloting participants to make sentences involving the past tense (e.g., Did Tom give me (something)?), which entailed conjugation of the *give* verbs. This seemed to focus their attention on correctly conjugating the verbs rather than on using the correct verbs according to the pragmatic rules which were the intended learning target. Fifth, piloting participants sometimes used the verb *morau*, which means ‘to receive’, instead of using *kureru* (or *kudasaru*). This
occurred because morau or kureru can be used to describe the same giving and receiving situation as shown in Example 8.

Example 8. Examples of morau and kureru piloting task sentences

Watashi-wa Tomu-ni morai-mashi-ta-ka
Self-NOM Tom-DAT receive-POLITE-PAST-QUESTION
Did I receive from Tom?

Tomu-wa watashi-ni kure-mashi-ta-ka
Tom-NOM self-DAT give-POLITE-PAST-QUESTION
Did Tom give me?

This reduced the opportunities for them to use kureru (or kudasaru) and to receive feedback if they mistakenly used ageru (or sashiageru) instead of kureru.

In order to address these problems with the piloting, the treatment tasks were modified in the following ways. First, the detailed situation presented prior to task performance was eliminated, and the tasks were made less artificial and clearer to participants by providing more instructions for each task item. Second, instead of including three tasks in the treatment session, it was decided to use only two, and the tasks focused only on ageru and kureru. Sashiageru and kudasaru were eliminated. Third, the new tasks did not require participants to provide reasons for their choices due to the time constraints. Fourth, the contexts were described in such a way that the future tense would be appropriate. In Japanese, verb conjugation is not necessary to describe future actions; the present and future tenses are morphologically the same, and if the communicative context does not disambiguate, a lexical item (e.g., ‘tomorrow’) can be used. Fifth, for each item in which participants had to create a sentence with a give verb,
the subject of the sentence was provided on the task sheet so that participants could not use *morau*. The modified versions of the treatment tasks are described below.

*Treatment tasks.* For the actual treatment session, two interaction tasks were created in the form of matching activities which only elicited the *give* verbs *ageru* and *kureru*. For Task One, participants received two separate materials: 1) an activity sheet and 2) a list of gifts and amounts of money to give. On the activity sheet, there were 16 items requiring participants to describe situations in which either the speaker was giving something to another person or someone was giving something to the speaker. For each task item, a visual illustration and description of the context was provided and the participant was asked to create a sentence with a *give* verb (see Example 9).

**Example 9. Examples of items in the treatment task**

a. 感冒（Saito） 

東京ディズニーランドのチケット (Tokyo Disneyland’s ticket)

How much will Saito-san give you in return?  
さいとうさんは私に___.

b. みやけ (Miyake)

$20

Which gift will you give to Miyake-san in return?  
私はみやけさんに__________.
As Example 9a shows, the pictures provided a context in which the participant was described as having given a ticket to Tokyo Disneyland to a man called Saito. Then the question “How much will Saito-san give you in return?” was asked. Below the question, an incomplete Japanese sentence, “さいとうさんは私に____” (“Saito-HON-TOP self-DAT _____”), was provided, creating an obligatory context for the participant to produce a sentence using a give verb, either ageru or kureru. From a list of money amounts on the sheet of another material, the participant had to decide which amount would be equivalent to the price of the ticket and create a sentence utilizing their selection. For Example 9b, the context involved a woman named Miyake giving the participant 20 dollars, the question asked was “Which gift will you give to Miyake-san in return?”, and the incomplete sentence provided was “私はみやけさんに__” (“Self-TOP Miyake-HON-DAT____”). From a list of possible gifts on another sheet, the participant had to decide which gift would be equivalent to the money received and create a sentence using that selection. In other words, participants had to supply the item he or she selected from the list, the accusative particle o to mark the object, and a verb to complete the sentence. It is important to point out that for Task One all the sentences involved ‘self’ (in-group), and acquaintances (out-group).

In Task One, there were eight sentences for which ageru was correct (ageru sentences) and another eight sentences for which kureru was correct (kureru sentences). Half of each type required participants to choose a money amount, while the other half of each type required a gift choice. The people appearing in the task were all from the participants’ age group. Besides the speaker (‘self’), half were men and the other half
were women. The gender of the people appearing in the tasks was counterbalanced across
1) the ageru and kureru sentences and 2) the money and gift items. All these
counterbalancing efforts were designed to avoid participants associating ageru and
kureru with an incorrect rule (e.g., related to gender or type of gift).

Task Two was essentially the same as Task One with one change. For this task, in
addition to the speaker (or ‘self’), the speaker’s family members were included in the in-
group: three male members (a father, older brother, and younger brother) and three
female members (a mother, older sister, and younger sister). The inclusion of sentences
involving family members (in-group) and acquaintances (out-group) was intended to
facilitate participants’ noticing of the in-group/out-group concept associated with the
give verbs. For Task Two, in addition to the counterbalancing measures taken in Task
One, the ages of the family members were counterbalanced in relation to ageru and
kureru. The other counterbalancing measures taken in Task One were also taken in Task
Two. During the treatment session, Task One was implemented first, followed by Task
Two.

These modified treatment tasks were pilot-tested with six students from the same
level of Japanese classes as the actual participants of the study. The observations and
comments from those piloting participants did not reveal any deficiencies with the tasks.
The tasks were also tested with 1) a native speaker of Japanese who was a graduate
student in linguistics and had taught Japanese for five years either as an instructor or a
teaching assistant in universities and 2) a near-native speaker of Japanese who had
studied Japanese for four years in college and worked in Japan for three years. The data
from the native and near-native speakers of Japanese showed that usage of *ageru* and *kureru* in the second pilot test matched what was intended 100 percent. In other words, for the sentences designed to elicit *ageru*, they used *ageru* without fail, and for the sentences designed to elicit *kureru*, they used *kureru* without fail. The complete version of treatment tasks used in the study is provided in Appendix I.

*Strategic planning instructions.* Upon the finalization of the treatment tasks, instructions for strategic planning were created and pilot-tested with five Japanese learners who were at the same level as the actual research participants to determine whether the instructions were easy to follow and whether the task design allowed participants adequate time to plan each task. In the instructions, participants were asked to prepare for the upcoming activity and to use the time to think about 1) making appropriate choices regarding the gifts and money amounts, 2) communicating naturally and fluently, and 3) using grammatically correct language. The instructions were neither detailed nor did they solely target either content or form as done in some previous strategic task-planning studies. Since clear benefits from detailed and targeted instructions were not suggested by the results of previous studies, and due to the fact that a very limited number of studies have investigated the connection between strategic planning and development, more general instructions seemed better suited. In addition, because this dissertation used think-alouds to determine if learners notice the target form during strategic planning, having fairly general planning instructions appeared to be sensible.
The comments from the pilot participants showed that the instructions were clear and that 10 minutes offered adequate time to plan each task. The amount of time they spent on planning each task is described in Table 1.

Table 1. Time pilot participants used for strategic planning

<table>
<thead>
<tr>
<th></th>
<th>Task 1</th>
<th>Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>9 min 35 sec</td>
<td>9 min 09 sec</td>
</tr>
<tr>
<td>Participant 2</td>
<td>7 min 52 sec</td>
<td>10 min 02 sec</td>
</tr>
<tr>
<td>Participant 3</td>
<td>10 min 12 sec</td>
<td>Not collected</td>
</tr>
<tr>
<td>Participant 4</td>
<td>9 min 50 sec</td>
<td>Not collected</td>
</tr>
<tr>
<td>Participant 5</td>
<td>19 min 44 sec</td>
<td>16 min 30 sec</td>
</tr>
</tbody>
</table>

After reviewing the recordings of the pilot participants’ think-alouds during the planning activity, it became clear that Participants 1 through 4 planned the content of the task, whereas Participant 5 not only planned for its content, but also rehearsed the entire task after planning the content. The aim of strategic planning is to plan the content of the task, which may involve some rehearsing of the task, but not to rehearse the entire task (otherwise, it would categorically be considered other types of pre-task strategy; see Ellis, 2005). Therefore, based on the data from pilot Participants 1 to 4, it was decided that 10 minutes was sufficient, and the allotted time was noted in the instructions. The strategic planning instructions are provided in Appendix J.
Pretests and Posttests

Pretests. Three assessment tasks were developed to evaluate participants’ L2 development of give verbs and their pragmatic rules prior to the treatment. Participants who did not demonstrate the ability to use ageru or used kureru in appropriate situations were eliminated from the analysis. The rationale for using multiple assessments came from Révész (2012), who found that the observed effectiveness of recasts depended on the type of outcome measures used. The present study’s assessments were 1) an oral task, 2) a written fill-in-the-blank production task, and 3) a written fill-in-the-blank recognition task.

Oral pretest. The oral task was similar in format to that of the treatment tasks with a few changes. For this task, participants received 1) an activity sheet and 2) a list of items to give and to lend. The pretest oral task consisted of 12 items. Half of the items (n=6) were designed to elicit give verbs: three for ageru, and three for kureru. The other half (n=6) were distractor items which were designed to elicit kasu, which means ‘to lend’. Kasu was chosen on the basis that 1) because it also requires a dative construction and was expected to work well in a matching activity. Of the six ageru and kureru items, two involved self (in-group) and an acquaintance (out-group), two involved a family member (in-group) and an acquaintance (out-group), and two involved a friend (in-group) and a celebrity⁷ (out-group) (these items were randomly presented). The context of the two items which involved a giving action between a friend and a celebrity was not

⁷ For the piloting of treatment tasks, celebrities were set to be the participant’s friends (in-group people). Since the setting appeared rather confusing for the piloting participants, the use of celebrities was eliminated from the actual treatment tasks. Celebrities were, however, used for the pre- and posttests tasks. For those tasks, celebrities were used as out-group people (someone he/she is not close to), not in-group people. This latter setting is obviously more natural and closer to reality for the participants.
addressed in the treatment tasks. These items were included to examine participants’
ability to generalize the pragmatic concept. Each of the two give verbs was tested in all
three contexts. Gender and age were counterbalanced across the three contexts and the
two give verbs.

For each item, participants were instructed to complete three steps: 1) to choose
an object from a list of items to give or to lend according to the question asked, 2) to
supply the dative particle *ni*, the accusative particle *o* to mark the object and 3) to
complete the sentence with a correct verb. Asking the participants to supply the particle
also functioned as a distractor because learners of Japanese typically find it difficult to
use particles correctly. I assumed that some participants might think the purpose of the
tests was to assess their usage of particles in dative constructions, a task some learners
find difficult. Examples of items from the oral pretest are found in Example 10.

Example 10. Examples of items on the oral pretest

Example 10. Examples of items on the oral pretest

| a. | 私はトムさん（  ）______________________________。
What will you lend him?
かぜをひきました。I got a cold. |
b. Which gift will you give to Chiba-san in return?

Which gift will you give to Chiba-san in return?

As shown in Example 10, each item provided a context and a question. Participants were asked to complete the incomplete sentence (i.e., fill in the blank) to answer the question. In Example 10a, the sentence started with “私はトムさん” or “Self-TOP Tom-HON”.

Participants had to supply the missing dative particle *ni* on the indirect object (as hinted at by the parentheses following the recipient’s name), followed by the item the participant selected from the list, the accusative particle *o* to mark the object, and a verb. The complete oral pretest is provided in Appendix K.

**Written production pretest.** The materials used for the written fill-in-the-blank production task were: 1) PowerPoint slides with test items, and 2) the answer sheet. This test included 24 items in total. One item was presented on each PowerPoint slide. Six items were designed to elicit give verbs: three for *ageru*, and three for *kureru*. The other 18 items were distractors designed to elicit *kasu* (‘to lend’), *kiku* (‘to ask’), and *oshieru* (‘to teach’). Each distractor verb was used in six items in the production task. The distractor verbs were selected on the basis that they also use the dative particle. For the six *ageru* and *kureru* items, two involved self (in-group) and an acquaintance (out-group), two involved a family member (in-group) and an acquaintance (out-group), and two
involved a friend (in-group) and a celebrity (out-group). As on the oral pretest, the same types of counterbalancing were done as on the oral pretest.

Each item provided an English sentence and a corresponding Japanese translation which was missing a particle and a verb. The object of the verb was already provided. In this test, the correct particle to fill in the blank was sometimes the dative particle, *ni*, but other times it was the topic particle, *wa*, or the accusative particle, *o*. An example from the written production pretest is provided below.

Example 11. Example item from the written production pretest

Tanaka-san will give (my)\(^8\) sister a bike.
田中さんは妹に自転車(  )______。

In Example 11, the parentheses in the Japanese sentence indicate where a particle is needed (in this case, the accusative particle *o*), and the underlined blank indicates where a verb is needed (in this case, the verb *kureru*). The complete written production pretest can be found in Appendix L.

**Written recognition pretest.** The written recognition task was essentially the same as the written production task. The only difference was that the participants had a word bank available to them during this task. The materials for the written fill-in-the-blank recognition task were: 1) PowerPoint slides with test items (one item on each slide), and 2) the answer sheet with a word bank. With the word bank, instead of generating the correct answer themselves as in the written production task, participants simply had to select the word they thought was correct for each item. The word bank consisted of 10

---

\(^8\) In the English sentence, parentheses mean that the word is not included in the Japanese translation. In the Japanese sentence below, parentheses mean that a particle should be used to fill in the blank. The blank in the Japanese sentence is not supposed to be filled with “my”.
verbs and 5 particles. All of the verbs which were provided required the use of the dative particle \textit{ni}. Two of them were the \textit{give} verbs (\textit{ageru} and \textit{kureru}), three were verbs which were assumed to be known to the participants since they had already been taught in their classes (\textit{oshieru} ‘to teach’, \textit{kasu} ‘to lend’, \textit{kiku} ‘to ask’), and five were expected to be unknown to the participants (\textit{susmeru} ‘to recommend’, \textit{nokosu} ‘to leave’, \textit{itadaku} ‘to receive’, \textit{yudaneru} ‘to entrust’, \textit{shoukaisuru} ‘to introduce’). The reason for including these five additional words was to reduce the chance of the participants paying attention to \textit{kureru}. That is, since it was assumed that the participants would not know \textit{kureru} and would use \textit{ageru} in the sentences created for \textit{kureru}, it seemed important to have some other words which were not used at all for the test items so that they would not have any particular reason to pay attention to \textit{kureru}. As in the written production pretest, six items were designed to elicit \textit{give} verbs: three for \textit{ageru}, and three for \textit{kureru}. The other 18 items were distractors designed to elicit \textit{kasu} (‘to lend’), \textit{kiku} (‘to ask’), and \textit{oshieru} (‘to teach’). Each distractor verb was used in six items in the production task. For the six \textit{ageru} and \textit{kureru} items, two involved self (in-group) and an acquaintance (out-group), two involved a family member (in-group) and an acquaintance (out-group), and two involved a friend (in-group) and a celebrity (out-group).

All of the pretests were pilot-tested with three pilot participants and did not reveal any apparent problems. The pretests were also pilot-tested with the native and near-native speakers of Japanese and verified that their usage of \textit{ageru} and \textit{kureru} match what was intended in the items in tests tasks. The complete recognition pretest is provided in Appendix M.
Posttests. Three assessment tasks were developed to evaluate participants’ knowledge of the pragmatic rules for give verbs following the treatment. The three posttests followed the types and formats of the pretests: 1) an oral task, 2) a written fill-in-the-blank production task, and 3) a written fill-in-the-blank recognition task. The only change made on the posttests was the number of sentences designed to elicit give verbs.\(^9\)

Oral posttest. For the oral posttest, the total number of items was 18. As on the pretest, 6 were distractors, but the number of items designed to elicit give verbs was increased to 12, with six for ageru and six for kureru. Among the 6 items for each give verb, two involved self (in-group) and an acquaintance (out-group), two involved a family member (in-group) and an acquaintance (out-group), and two involved a friend (in-group) and a celebrity (out-group). As before, gender and age were counterbalanced across verbs and contexts. The distractors were used in order to deter the participants from forming the target rule during the test phase, not to hide the fact that give verbs were the focus of the study as in the pretests. The complete oral posttest is can be found in Appendix N.

Written production and recognition posttests. Each of the written posttests included 36 items in total; 18 were give verbs and 18 were distractors. The number of give verb items was increased compared to the pretests, but the number of distractors was the same as in the pretests. Among the 18 give verb sentences, nine were for ageru and

\(^9\) The number of sentences eliciting give verbs was smaller in pretests compare to posttests because the purpose of the pretests was to ensure that 1) the participants had knowledge of ageru and could use ageru and 2) the participants did not have knowledge of kureru, another give verb. For this purpose, it did not seem necessary to have larger numbers of items. The posttests needed larger numbers of items to perform statistical analysis between two experimental groups. In retrospect, however, it would probably have been better to have the same numbers of items on the pretests and posttests because a repeated measures ANOVA was used with the data from pretests and posttests.
nine were for kureru, and among each of these 9, three involved self (in-group) and an acquaintance (out-group), three involved a family member (in-group) and an acquaintance (out-group), and three involved a friend (in-group) and a celebrity (out-group). Gender and age were counterbalanced across verbs and contexts, and distractors were again used to deter the participants from forming the target rule during the test phase. The posttests were not pilot tested, since the pretests were the same in format and the pilot testing of the pretests had not revealed any problems. The written production and recognition posttests are provided in Appendices O and P, respectively.

Measures of Noticing

Along with the data from the treatment task interactions, three introspective measures were used in this study to gauge participants’ noticing of the pragmatic concept associated with give verbs: think-aloud protocols, stimulated recall interviews, and an exit questionnaire.

Think-aloud protocols. Think-aloud protocols were used during the strategic pre-task planning in order to record what the participants reported being noticed while planning for the treatment task. For the present study, participants were not asked to provide explanations or justifications of their thoughts, since verbalizations of that type have been argued to be reactive on cognitive processes (Ericsson & Simon, 1993; Leow & Morgan-Short, 2004). The participants were instead instructed simply to verbalize whatever thoughts naturally came to mind.

Stimulated recall interviews. Another introspective measure, the stimulated recall interview, was used to capture participants’ noticing of the linguistic target during the
treatment tasks. A stimulated recall interview was performed one-on-one immediately after each participant’s treatment session in order to reduce the influence of memory decay on the data. This stimulated recall was conducted by a research assistant who was a graduate student in linguistics and a teaching assistant of Japanese at Georgetown University. She had extensive experience conducting interviews in sociolinguistics research.

In order to train the research assistant, the researcher first defined stimulated recall and explained the rationale for using it for the present research. Then, the researcher demonstrated how to conduct the interview by carrying out a role-play with the research assistant using a video recording of a pilot participant engaging in one of the treatment tasks (the research assistant pretended to be the participant during this role play). The researcher and the research assistant then switched roles, and the research assistant practiced conducting a stimulated recall interview with the researcher playing the role of the participant. Several important guidelines were mentioned to the research assistant: 1) to ensure that the visual images and sounds were clear before conducting the interview, 2) to state clearly that participants must report only what they had been thinking at the time of the original task performance, 3) to refrain from asking for reasons or explanations during the interviews, and 4) to act as a passive listener and to demonstrate that social interaction was not intended during the interviews.

The primary reason for using a research assistant for the stimulated recall interviews was to collect more valid data. It was assumed that because the researcher had engaged in the treatment tasks with the participants, some of them might feel awkward
engaging in the stimulated recall interview with the same researcher and might even refrain from providing their true thoughts. With the researcher not participating in the interview, it was expected that the participants would feel more comfortable making comments.

*Exit questionnaire.* Another measure, an exit questionnaire, was triangulated with the data obtained from think-alouds and stimulated recalls. Six questions were designed to assess participants’ noticing of the pragmatic rules: 1) Do you have any other comments about the activities?, 2) What do you think was the focus of this research project?, 3) Is there anything you noticed while you were doing the activities? 4) Is there anything you noticed about what the Japanese speaker was saying to you?, 5) Is there anything you noticed about the Japanese language?, 6) If you had to explain the verbs *ageru* and *kureru* to a friend who didn’t know about them, what might you say (please include any ideas you have, even if they’re vague and you’re not sure)? These questions were arranged in a manner that became progressively more specific. In other words, at the beginning, I gave participants a chance to report whatever came to mind without being primed. The last question was provided on a separate sheet from the rest of the questions, and it was given after all the previous questions had been answered (see Appendix Q).

*Perceived Difficulty and Helpfulness Measures*

In the exit questionnaire, there was a question regarding the perceived difficulty of the treatment tasks, and (or the participants who engaged in strategic planning) a question on the perceived helpfulness of planning. For the latter two questions, the
participants were asked to provide ratings on a 5-point Likert scale along with an explanation of their ratings.

Coding of Noticing

Data from the participants in the stimulated recall groups (+P/SR and -P/SR) were coded in order to address the research question about whether strategic planning has an effect on noticing of L2 forms and contextual factors relevant to pragmatics. Four data sources were used: 1) the verbal reports from the think-aloud protocols during planning (only from the +P/SR group\textsuperscript{10}.), 2) the verbal reports from the stimulated recall interview, 3) the interactions during the two treatment tasks, and 4) the reports from the exit questionnaire. All the data were coded either notice (to indicate that the participants had at least demonstrated awareness of the linguistic target at the level of noticing) or no report. In discussing L2 pragmatic development, Schmidt (1993) stated that learners “must notice both the form and the relevant contextual factors…if it is to be intake for learning” (p. 34). Therefore, the coding system developed for this analysis focused on both form and contextual factors. In addition, after reviewing the four different data sources, it became apparent that they produced different types of data and that more than one coding schema was necessary. The different coding schemata developed for coding the data from the four data sources for both form and contextual factors are described separately below (see also Table 2).

\textsuperscript{10} For this first data source, I could only code the +P/SR group. This is because the -P/SR group did not have strategic planning; therefore, there is no-planning data from the latter group.
Noticing of Form

Since the participants already knew the word ageru, the form which they needed to notice was kureru. For the verbal protocols from the think-alouds and stimulated recalls, a participant was coded noticing of form when he/she 1) mentioned kureru (or kuremasu, the polite form of kureru), 2) asked about kureru (or kuremasu) or 3) mentioned mixing up relevant verbs or getting corrected on their relevant verb usage. The examples below were taken from data from the present study.

Example 12. Noticing of form in think-alouds and stimulated recall protocols

Mentioned kureru

a. P20\(^{11}\) [think-alouds]: ...movie ticket is 20 dollars....kuremasu, kuremasu and then 60 dollars is, um...I guess radio-cassette player.

Asked about kureru

a. P29 [stimulated recalls]: Why it’s kureru?

Mentioned mixing up relevant verbs or being corrected on relevant verb usage

a. P22 [stimulated recalls]: I was thinking that I forgot the choice that I was planning to say and about the verb confusion which one should I use.

b. P23 [stimulated recalls]: What was I thinking then? I was confused, I thought it was just agemasu, but then sensee [‘teacher’] said kuremasu, corrected me, so I was getting confused from that point, yeah.

From the interaction data, a participant was coded noticing of form when he/she 1) repeated kureru (or kuremasu) after a recast from the researcher, 2) asked about kureru

\(^{11}\) P20 stands for participant 20. Each participant was given number.
(or kuremasu), or 3) used kureru (or kuremasu) in an answer. Based on the questionnaire comments (excluding question 6, which asked, “If you had to explain the verbs ageru and kureru to a friend who didn’t know about them, what might you say?”), each participant was coded noticing of form when he/she 1) wrote kureru (or kuremasu), 2) wrote that there are two verbs for to give, 3) wrote about mixing up the relevant verbs or being corrected on the relevant verbs.

Example 13. Noticing of form on questionnaires

Wrote there are two verbs for to give

a. P28: There are many ways to say the same thing. For instance, there are two different ways to say ‘to give’.

Wrote about mixing up the relevant verbs or being corrected on the relevant verbs

a. P18: She corrected me when I used the wrong verb.

In sum, when a participant demonstrated one of the above types of report in any of the four sources of data, he/she was coded as noticing of form. When a participant demonstrated none of the above, he/she was coded as no report.

Noticing of Contextual Factors

The relevant contextual factors related to the pragmatic rules of the give verbs presented through the treatment tasks were: 1) self and acquaintances (or self and non-self) and 2) family members and acquaintances (or family and non-family). The participants were exposed to these two contextual factors in order to understand the in-group and out-group principle related to the two give verbs. Based on the think-alouds, stimulated recalls, interactions, and questionnaire, a participant was coded noticing of self.
and/or non-self when the participant explicitly mentioned self and/or non-self (or other) along with one of the verbs (ageru, kureru). A participant was coded noticing of family and non-family when he/she explicitly mentioned family. A participant was coded noticing of in-group and out-group when he/she mentioned designations from which the concept could be inferred, for example, “people who are close”, “people related”, and “immediate group”. The participants’ uses of words such as watashi (“I”, “me”, or “my”) in their actual answers or repetitions of answers during the task interactions were not included in this coding; in order to be classified as noticing of contextual factors, they had to mention the contextual factors explicitly.

Example 14. Noticing of contextual factors in the think-alouds, stimulated recall protocols, interactions, and questionnaire

Noticing of self and/or non-self


b. P23 [stimulated recalls]: oh and I guess I was thinking kuremasu when there was watashi [‘I’], yeah so I was using kuremasu

c. P30: [questionnaire]: I thought ageru was me → another person and kureru was another person → me.

Noticing of family and non-family

a. P29 [questionnaire]: I guessed family members but I don’t know.
b. P25 [questionnaire]: You use kureru when it is a family member giving
the gift to someone else.

Noticing of in-group and out-group

a. P16 [stimulated recalls]: I was.. still trying to figure out the prices used
for the new sheet and.. ah. that’s about it, that and also still trying to
figure out the differences between agemasu and kuremasu, between the
two, is it . differ.. to me is kuremasu and from me is agemasu, or would
that be to family members, someone close to you or I couldn’t figure it
out.\textsuperscript{12}

b. P16 [questionnaire]: I’m not sure but ageru is to give from yourself or
those close to you and kureru is to give from someone not close to you.\textsuperscript{13}

Each participant was coded in four ways for contextual factors: 1) no report, 2)
noticing of self and/or non-self, 3) noticing of family and non-family, and 4) noticing of
in-group and out-group. Although each data source was coded independently,
classifications were made at the level of participant. In other words, if a participant was
coded noticing of self and/or non-self in one of the data sources, he/she was even in the
absence of such coding in the other data sources. It was only if none of the above was
demonstrated in any of the data sources that a participant was coded no report.

In addition, noticing of in-group and out-group was counted as “complete”
noticing of the pragmatic rules. Noticing of self and/or non-self and noticing of family

\textsuperscript{12} This comment was also counted as noticing of self and/or non-self and noticing of family and non-family.

\textsuperscript{13} This comment was also counted as noticing of self and/or non-self.
and non-family was counted as partial if the participant only coded one of them. However, when the participant was coded for both noticing of self and/or non-self and noticing of family and non-family the occurrence was counted as exhibiting complete noticing of the pragmatic concept.

Table 2. Coding scheme of noticing

<table>
<thead>
<tr>
<th>Data source</th>
<th>Noticing of form</th>
<th>Noticing of contextual factors</th>
</tr>
</thead>
</table>
| Verbal reports from think-alouds and stimulated recalls | a. Mentioned *kureru*,  
b. Asked about *kureru*, or  
c. Mentioned mixing up *give* verbs or being corrected on the verb usage | a. **Self and/or non-self:** Mentioned self and/or non-self along with one of the *give* verbs besides using these words in their answers or repetitions of answers  
b. **Family and non-family:** Mentioned family  
c. **In-group and out-group:** Mentioned relationships from which an in group/out-group distinction can be inferred (e.g., “people who are close”, “people related”, “immediate group”) |
| Interaction                   | a. Repeated *kureru* after researcher’s recast,  
b. Asked about *kureru*, or  
c. Used *kureru* in an answer | The same coding scheme used for the think-alouds and stimulated recalls |
| Exit questionnaire            | a. Wrote *kureru*,  
b. Wrote there are two verbs for *to give*, or  
c. Wrote about mixing up *give* verbs or being corrected on the verbs | The same coding scheme used for the think-alouds and stimulated recalls |
Inter-coder Reliability

The audio- and video-recorded data from 1) the think-aloud protocols during planning, 2) the stimulated recall interviews, and 3) the interactions during the two treatment tasks were transcribed by the researcher for coding purposes. Then, 25% of the data from the stimulated recall groups (i.e., 5 of the 20 participants) from the four data sources (think-alouds, stimulated recalls, task interactions, and questionnaires) was randomly selected to be coded by both the researcher and the research assistant using the coding schema described earlier. The participants in the -P/SR group did not engage in planning; therefore, no think-aloud data were collected or coded for those participants.

An inter-coder reliability analysis using the Kappa statistic was performed to determine consistency between the coders. When all the data sources were combined together, kappa coefficient was $\kappa = 1.00$ ($p < .001$). How each of the coder’s coding data is found in Table 3.
Table 3. Coding from the two coders on noticing of form

<table>
<thead>
<tr>
<th>Participant 1 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
<tr>
<td>Research assistant</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 2 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
<tr>
<td>Research assistant</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 3 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>No report</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
<tr>
<td>Research assistant</td>
<td>No report</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 4 (-P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>N/A</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
<tr>
<td>Research assistant</td>
<td>N/A</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 5 (-P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>N/A</td>
<td>No report</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
<tr>
<td>Research assistant</td>
<td>N/A</td>
<td>No report</td>
<td>Notice</td>
<td>Notice</td>
<td>Notice</td>
</tr>
</tbody>
</table>

As for the noticing of contextual factors, for all of the data sources together, kappa coefficient was $\kappa = 0.90 \ (p < .001)$. Each of coder’s coding is found in Table 3. The disagreement in the coding was discussed by the coders and they agreed to keep the researcher’s coding. Because of the sufficiently high level of agreements, the rest of the data was coded only by the researcher before being submitted to data analysis.
Table 4. Coding from the two coders on noticing of contextual factors

<table>
<thead>
<tr>
<th>Participant 1 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>Self</td>
<td>Self</td>
<td>No report</td>
<td>Self Family</td>
<td></td>
</tr>
<tr>
<td>Research assistant</td>
<td>Self</td>
<td>Self</td>
<td>No report</td>
<td>Self Family</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 2 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>No report</td>
<td>Self</td>
<td>No report</td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>Research assistant</td>
<td>No report</td>
<td>Self</td>
<td>No report</td>
<td>Self</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 3 (+P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td></td>
</tr>
<tr>
<td>Research assistant</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 4 (-P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>N/A</td>
<td>Self Family</td>
<td>No report</td>
<td>Family</td>
<td></td>
</tr>
<tr>
<td>Research assistant</td>
<td>N/A</td>
<td>Family</td>
<td>No report</td>
<td>Family</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant 5 (-P/SR)</th>
<th>Coder</th>
<th>Think-alouds</th>
<th>Stimulated recall</th>
<th>Task interaction</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researcher</td>
<td>N/A</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td></td>
</tr>
<tr>
<td>Research assistant</td>
<td>N/A</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td></td>
</tr>
</tbody>
</table>

Scoring of Posttests

The data from the oral and two written tests were used to gauge participants’ development of the linguistic target. For each of the *give* verb items, on all the tests, a correct answer was awarded one point; an incorrect answer received zero points. This means that when a participant used *ageru* for a *kureru* item or *kureru* for an *ageru* item, the answer was given zero points.

Before deciding on the scoring system above, I considered awarding partial credit when a participant answered with a *kureru* for an *ageru* item because I thought using *kureru*, even incorrectly, may show some destabilization of their interlanguage and it may
deserve some credit. When I closely examined each participant’s task interaction during the treatment session, however, it became clear that all but one participant used *kureru* instead of *ageru* for a sentence that involved giving actions. I interpreted this to be some evidence that almost all the participants at least knew that *kureru* is relevant to use. In addition, using *kureru* itself does not necessary show that participants are closer to the acquisition of target form. For example, while one participant who used *kureru* three times in the oral posttest commented in his exit questionnaire\(^{14}\) that “…when someone gives something to me, then I can use *kureru*”, another participant who used *kureru* four times commented on the exit questionnaire that “I could not figure out when to use which for the life of me. I tried to go by price of the object being given, but that didn’t quite work out, so I started guessing”. All and all, it seemed more logical to use the measure which showed accurate use of the two *give* verbs. In next chapter, Results, I will describe the usage of the two verbs and how each verb was used accurately in posttests.

\(^{14}\) I am aware that participant’s comment on the exit questionnaire cannot be equated with cognitive processes during the treatment and posttests. However, the *possibility* that the comment represented his or her cognitive process during treatment and posttests cannot be excluded; therefore, I am using the comment to validate my point.
Chapter 3: Results

In this dissertation, the following questions were addressed: 1) do learners who receive recasts following strategic planning show greater L2 development than learners who have no planning time available? and 2) do learners who receive recasts following strategic planning show greater noticing of L2 linguistic forms and pragmatic factors than learners who have no planning time available? The first question, related to L2 development, also addressed two sub-questions due to the design of this present experiment: (a) are there any differences in posttest performances according to verb? (b) are there any differences in posttest performances according to the various facets of the pragmatic principle? These sub-questions were necessary because first, the participants were already familiar with *ageru* prior to experiment but not familiar with *kureru*; and second, it was possible that certain aspects of the pragmatic principle may be learned more readily than others, and there may be an interaction with planning conditions. For all statistical tests, the alpha level was set at p < .05, effect sizes are reported as eta-squared values ($\eta^2$). Following Cohen (1977), for this experiment, eta-squared values were interpreted as small if they were between 0.02-0.13, as medium if they were between 0.13-0.26, and as large if they were greater than 0.26.

**Research Question 1:**

*Do Learners Who Receive Recasts Following Strategic Planning Show Greater L2 Development Than Learners Who Have No Planning Time Available?*

To address this question, the posttest results from the strategic planning and posttest group (+P/T) and the no strategic planning and posttest group (-P/T) were
compared. The remaining treatment groups, the strategic planning and stimulated recall group (+P/SR) and the no strategic planning and stimulated recall group (-P/SR), were not administered any posttests and therefore did not produce data relevant to answering this question. In reporting the results for Research Question 1, the strategic planning and posttest group (+P/T) and no strategic planning and posttest group (-P/T) will be called the Planning Group and the No Planning Group, respectively. Overall results are first described first and then the sub-questions are addressed.

**Overall Results**

**Pretests.** Descriptive statistics from the Planning Group and the No Planning Group in each of the pretests (oral, written production, or written recognition) are provided in Table 5. The numbers in the table are in percentage format instead of numbers of correct answers. This is due to the fact that numbers of items in the pretests and those in the posttests are different; therefore, in order to observe the changes between the pretests and the posttests it was necessary to convert the number of correct items to percentages.

### Table 5. Pretests accuracy by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Oral</th>
<th>Written production</th>
<th>Written recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Planning</td>
<td>48.81</td>
<td>4.46</td>
<td>50.00</td>
</tr>
<tr>
<td>No Planning</td>
<td>47.92</td>
<td>5.69</td>
<td>48.96</td>
</tr>
</tbody>
</table>
Independent samples \( t \)-tests revealed no significant difference in oral pretest accuracy (\( t = 0.47, p = 0.64 \)), in written production pretest scores (\( t = 0.93, p = 0.36 \)), and in written recognition pretest scores (\( t = 0.10, p = 0.93 \)). Therefore, there is no significant difference between the two groups in their accurate usage of give verbs.

**Oral tests.** A repeated measures ANOVA was performed to evaluate whether there was a significant accuracy difference between the pretest and posttest and whether a significant difference existed between the two groups. The means and standard deviations of the accuracy from the Planning Group and the No Planning Group on the oral pretest and posttest are provided in Table 6.

Table 6. Oral pretest and posttest accuracy by group

<table>
<thead>
<tr>
<th>Group</th>
<th>Accuracy</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( M )</td>
<td>( M )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( SD )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Planning</td>
<td>48.81</td>
<td>4.46</td>
<td>51.78</td>
</tr>
<tr>
<td>No Planning</td>
<td>47.91</td>
<td>5.69</td>
<td>62.50</td>
</tr>
</tbody>
</table>

The statistical analysis revealed a significant difference between the pretest and the posttest (\( F = 11.57, p = 0.003, \eta^2 = 0.29 \)), indicating that the participants were significantly more accurate in the posttest compare to the pretest. There was also a significant interaction between the time (pretest, posttest) and the groups (\( F = 5.06, p = 0.03, \eta^2 = 0.15 \)). Because the two groups were not significantly different at the point of the pretest, this indicates the two groups were significantly different in terms of the rate
of improvement in accuracy. This, therefore, indicates that the No Planning Group demonstrated a significantly higher rate of improvement compared to the Planning Group (see Figure 4).

Figure 4. Mean percentage of correct answers in oral pretest and posttest by group

Written production tests. The descriptive statistics from the written production pretest and posttest are provided in Table 7.
Table 7. Written production pretest and posttest accuracy by group

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Planning</td>
<td>50.00</td>
<td>0.00</td>
</tr>
<tr>
<td>No Planning</td>
<td>48.96</td>
<td>4.17</td>
</tr>
</tbody>
</table>

The statistical analysis revealed a significant difference between the pretest and the posttest ($F = 11.50$, $p = 0.002$, $\eta^2 = 0.29$), indicating that the participants were significantly more accurate in the posttest compared to the pretest. Interaction between the time (pretest, posttest) and the groups was not significant ($F = 1.32$, $p = 0.26$, $\eta^2 = 0.05$). This indicates the two groups were not significantly different in terms of the rate of improvement in accuracy (see Figure 5).
Figure 5. Mean percentage of correct answers in written production pretest and posttest by group.

*Written recognition tests.* The descriptive statistics from the written recognition posttest are provided in Table 8.
Table 8. Written recognition pretest and posttest accuracy by group

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Planning</td>
<td>48.81</td>
<td>4.46</td>
</tr>
<tr>
<td>No Planning</td>
<td>48.96</td>
<td>4.17</td>
</tr>
</tbody>
</table>

The statistical analysis revealed a significant difference between the pretest and the posttest ($F = 12.23, p = 0.002, \eta^2 = 0.30$), indicating that the participants were significantly more accurate in the posttest compared to the pretest. Interaction between the time (pretest, posttest) and the groups was not significant ($F = .24, p = 0.63, \eta^2 = 0.01$). This indicates the two groups were not significantly different in terms of the rate of improvement in accuracy (see Figure 6).
Figure 6. Mean percentage of correct answers in written recognition pretest and posttest by group

Summary of three tests results. The statistical analysis revealed the participants were significantly more accurate in the posttest compare to the pretest in all tests. In oral tests, results indicated the No Planning Group improved at a significantly higher rate than the Planning Group. No difference between the groups was, however, found in the other two tests.
Ageru and Kureru Posttest Results

Oral Posttest. A repeated-measures ANOVA was used to test if there was a test score difference between the two verbs (i.e., *ageru* test items and *kureru* test items), and if there was an interaction between the groups and the verbs (for example, if one group performed better than the other on *ageru* items but performed worse than the other group on *kureru* items). Then, an independent samples *t*-test was performed for each verb to evaluate whether the two groups’ scores differed significantly from one another. The means and standard deviation of the scores for the two verbs from the Planning Group and the No Planning Group on the oral posttest are provided in Table 9. On the oral posttest, the total number of items for each verb was six; therefore, the maximum score participants could obtain for each verb was six.

Table 9. Oral posttest scores for each verb by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Ageru (n = 6)</th>
<th>Kureru (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>M</em></td>
<td><em>SD</em></td>
</tr>
<tr>
<td>Planning</td>
<td>5.07</td>
<td>1.27</td>
</tr>
<tr>
<td>No Planning</td>
<td>5.13</td>
<td>1.15</td>
</tr>
</tbody>
</table>

The repeated-measures ANOVA revealed a significant difference between the verbs ($F = 53.61, p < 0.01, \eta^2 = 0.66$). Participants scored significantly higher on the *ageru* items compared to the *kureru* items. The interaction between the planning conditions and the verbs was not significant ($F = 1.67, p = 0.21, \eta^2 = 0.06$). Figure 7 illustrates each group’s score on each of the two verbs.
In order to determine if there was a significant test score difference between the two groups for each verb, a series of independent samples $t$-tests was performed. The $t$-test revealed no significant difference between the two groups for either the *ageru* items ($t = -0.12, p = 0.90, \eta^2 = 0.001$) or the *kureru* items ($t = -2.02, p = 0.053, \eta^2 = 0.13$).

However, it must be noted that the difference in the groups’ scores on the *kureru* items approached significance.

Figure 7. Mean oral posttest score for each verb by group
Written production posttest. The descriptive statistics from the written production posttest are provided in Table 10. There were nine items for each verb in the production test.

Table 10. Written production posttest scores for each verb by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Ageru (n=9)</th>
<th>Kureru (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Planning</td>
<td>9.00</td>
<td>0.00$^{15}$</td>
</tr>
<tr>
<td>No Planning</td>
<td>8.44</td>
<td>1.09</td>
</tr>
</tbody>
</table>

The results from a repeated-measures ANOVA showed there was a significant difference between the two verbs ($F = 185.47, p < 0.01, \eta^2 = 0.87$), again indicating that participants performed significantly better on the *ageru* items than on the *kureru* items. The interaction between the planning conditions and the verbs was significant ($F = 4.45, p = 0.04, \eta^2 = 0.14$). Figure 8 depicts the each group’s score on each of the two verbs.

---

$^{15}$ The mean is 9 and the SD is 0 indicates everyone in The Planning Group only used *ageru* on the *ageru* items
A series of independent samples $t$-tests was run to determine whether the difference between the Planning Group and the No Planning Group in the number of accurate answers for each verb was significant. Independent samples $t$-tests revealed no significant difference between the two groups for the *ageru* items ($t = 1.92, p = 0.07, \eta^2 = 0.12$) or for the *kureru* items ($t = -1.59, p = 0.12, \eta^2 = 0.08$). This set of $t$-tests results indicate that the significant interaction found earlier might have been because the directions of the differences were different according to verb.
Written recognition posttest. Table 11 shows the mean scores and standard deviations of the correct responses on the two verbs from the Planning Group and the No Planning Group on the written recognition test. The number of items for each verb was nine.

Table 11. Written recognition posttest scores for each verb by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Ageru (n = 9)</th>
<th>Kureru (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Planning</td>
<td>8.64</td>
<td>1.08</td>
</tr>
<tr>
<td>No Planning</td>
<td>8.12</td>
<td>1.63</td>
</tr>
</tbody>
</table>

Results from a repeated-measures ANOVA revealed the difference in test scores according to the verb was significant \( (F = 81.84, p < 0.01, \eta^2 = 0.75) \), indicating that the test scores for the *ageru* items were significantly higher than those for the *kureru* items. The interaction between the groups and the two verbs was not significant \( (F = 1.65, p = 0.21, \eta^2 = 0.06) \). Figure 9 illustrates each group’s score for each of the two verbs.
Independent samples $t$-tests were performed to determine the significance of the differences between the Planning Group and the No Planning Group in their scores for each verb. The results revealed no significant difference between the two groups for the *ageru* items ($t = 1.01, p = 0.32, \eta^2 = 0.04$) or the *kureru* items ($t = -1.00, p = 0.33, \eta^2 = 0.04$).

Ageru and kureru frequency of usage during posttests. As another way to investigate *ageru* and *kureru*, the participants’ frequency of usage (whether accurate or
inaccurate) of the two verbs, *ageru* and *kureru*, during the posttests was analyzed. First of all, a simple count of the numbers of participants in each treatment group, the Planning Group and the No Planning Group, who only used *ageru* during each posttest is described in Table 12.

Table 12. Numbers of participants who used only *ageru* during each posttest

<table>
<thead>
<tr>
<th>Type of test</th>
<th>Planning (N = 14)</th>
<th>No Planning (N = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>6 (43%)</td>
<td>3 (19%)</td>
</tr>
<tr>
<td>Written production</td>
<td>9 (64%)</td>
<td>3 (19%)</td>
</tr>
<tr>
<td>Written recognition</td>
<td>9 (64%)</td>
<td>1 (6%)</td>
</tr>
</tbody>
</table>

Table 12 shows that among the 14 participants in the Planning Group, nearly half (*n* = 6) in the oral test, and more than half (*n* = 9) in the written production and recognition tests, used only *ageru* (i.e., *kureru* did not appear in their answers). In contrast, among the 16 participants in the No Planning Group, only three participants used only *ageru* in both the oral and written production tests, and just one used exclusively *ageru* on the written recognition test.

Table 13 shows the mean frequencies and standard deviations of *ageru* and *kureru* use by the participants in the Planning Group and the No Planning Group for the three posttests.
Table 13. Frequencies of verb use in each posttest by group

<table>
<thead>
<tr>
<th>Type of tests</th>
<th>Ageru Planning</th>
<th>Ageru No Planning</th>
<th>Kureru Planning</th>
<th>Kureru No Planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>9.93</td>
<td>2.62</td>
<td>2.00</td>
<td>2.60</td>
</tr>
<tr>
<td>(n = 12)</td>
<td>8.75</td>
<td>2.44</td>
<td>3.25</td>
<td>2.44</td>
</tr>
<tr>
<td>Written production</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>16.86</td>
<td>2.45</td>
<td>1.14</td>
<td>2.45</td>
</tr>
<tr>
<td>(n = 18)</td>
<td>14.75</td>
<td>2.96</td>
<td>3.25</td>
<td>2.96</td>
</tr>
<tr>
<td>Written recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>15.57</td>
<td>3.92</td>
<td>2.43</td>
<td>3.92</td>
</tr>
<tr>
<td>(n = 18)</td>
<td>13.81</td>
<td>2.95</td>
<td>4.06</td>
<td>3.04</td>
</tr>
</tbody>
</table>

A series of independent samples t-tests was run to determine if the difference in the frequency of usage of *ageru* and *kureru* between the Planning Group and the No Planning Group reached significance for each posttest. The results for *ageru* show that on the oral test ($t = 1.28$, $p = 0.21$, $\eta^2 = 0.06$) and the recognition test ($t = 1.40$, $p = 0.17$, $\eta^2 = 0.07$) significance was not achieved. However, there was a statistically significant difference between the groups on the production test ($t = 2.11$, $p = 0.04$, $\eta^2 = 0.14$). Similarly, the results for *kureru* did not reveal a significant difference between the two groups on the oral test ($t = -1.36$, $p = 0.19$, $\eta^2 = 0.06$) or the written recognition test ($t = -1.28$, $p = 0.21$, $\eta^2 = 0.06$). However, the difference was statistically significant on the written production
test \((t = -2.11, p = 0.04, \eta^2 = 0.13)\). These results indicate that on the written production test, 1) the Planning Group used \(ageru\) significantly more often than the No Planning Group, and 2) the No Planning Group used \(kureru\) significantly more often than the Planning Group.

**Summary of \(ageru\) and \(kureru\) posttest results.** Statistical analyses revealed that for the scores on both \(ageru\) items and \(kureru\) items considered separately, there was no significant difference between the groups on any test. However, the results for the \(kureru\) items on the oral test showed a difference between the groups which approached statistical significance. As for the frequency of usage of the two verbs, the results from a simple count of the numbers of participants in each treatment group who only used \(ageru\) during each posttest revealed that the number was notably higher in the Planning Group compare to that of the No Planning Group. When the Planning Group and the No Planning Group were compared on their frequency of usage of the two verbs, differences reached statistical significance only for the written production test. The \(\eta^2\) values indicated a medium effect for \(ageru\), and small to medium effect for \(kureru\).

**Results from the Three Test-Item Types**

*Oral Posttest.* The 12 test items targeting \(give\) verbs in the oral test were evenly divided into three types related to different possible aspects of the pragmatic principle underlying the use of \(ageru\) versus \(kureru\): 1) Self and/or non-self \((n = 4)\), 2) family and non-family \((n = 4)\), and 3) in-group and out-group \((n = 4)\). Here, they will be referred as 1) Self, 2) Family, and 3) In/Out-group, respectively. The differences in the test scores
between the Planning Group and the No Planning Group for each type were also examined (see Table 14).

Table 14. Oral posttest scores of test item types by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Self</th>
<th></th>
<th>Family</th>
<th></th>
<th>In/Out-group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Planning</td>
<td>2.50</td>
<td>0.76</td>
<td>1.86</td>
<td>0.66</td>
<td>1.86</td>
<td>0.36</td>
</tr>
<tr>
<td>No Planning</td>
<td>2.75</td>
<td>1.00</td>
<td>2.44</td>
<td>1.03</td>
<td>2.31</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Independent samples t-tests revealed no significant difference between the groups for Self ($t = -0.76, p = 0.45, \eta^2 = 0.01$) or for Family ($t = -1.80, p = 0.08, \eta^2 = 0.10$).

However, a significant difference between the groups was found for the In/Out-group item type ($t = -2.90, p = 0.007, \eta^2 = 0.27$), indicating that the No Planning Group had a significantly higher score on the In/Out-group items compared to the Planning Group. In order to determine whether there was an overall difference (regardless of planning group conditions) between the item types, and whether the interaction between the planning group conditions and the item types was significant, a repeated-measures ANOVA with planning condition as the between-subjects factor and item types as the within-subjects factor was analyzed. The analysis revealed the difference in test scores according to item types was significant ($F = 5.12, p = 0.009, \eta^2 = 0.16$), but the interaction was not significant ($F = .408, p = 0.67, \eta^2 = 0.01$). Post hoc analysis reported a significant differences between Self and Family ($F = 6.23, p = 0.02, \eta^2 = 0.18$) also Self and In/Out-Group ($F = 9.70, p = 0.004, \eta^2 = 0.26$). No significance was found between Family and
In/Out-Group \((F = 0.11, p = 0.74, \eta^2 = 0.004)\). These *post hoc* results indicated that on the oral posttest the participants were significantly more accurate on Self items than on the other two types of items.

The results on the three item types were further tested separately by verb. The descriptive statistics are presented below in Table 15. Because the number of items for each of the three item types was four and this was further divided by the two verbs, for each cell, the maximum possible score was two. It must be mention that due to the small number of items, there is an increase in the risk of errors in statistical testing.

Table 15. Oral posttest scores for each verb by test-item types and group

<table>
<thead>
<tr>
<th>Verbs &amp; Groups</th>
<th>Self</th>
<th>Family</th>
<th>In/Out-group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
</tr>
<tr>
<td><strong>Ageru</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td>1.79</td>
<td>0.43</td>
<td>1.57</td>
</tr>
<tr>
<td>No planning</td>
<td>1.69</td>
<td>0.60</td>
<td>1.62</td>
</tr>
<tr>
<td><strong>Kureru</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning</td>
<td>0.71</td>
<td>0.91</td>
<td>0.29</td>
</tr>
<tr>
<td>No planning</td>
<td>1.06</td>
<td>0.85</td>
<td>0.81</td>
</tr>
</tbody>
</table>

According to independent samples *t*-tests, no significant differences were found for any of the comparisons between the Planning Group and the No Planning Group, whether for the *ageru* items focusing on Self \((t = 0.51, p = 0.62, \eta^2 = 0.01)\), Family \((t = -0.21, p = 0.83, \eta^2 = 0.001)\), or In/Out-group \((t = -0.62, p = 0.54, \eta^2 = 0.01)\), or for the *kureru* items
focusing on Self \( (t = -1.08, p = 0.29, \eta^2 = 0.04) \), Family \( (t = -1.95, p = 0.06, \eta^2 = 0.12) \), or In/Out-group \( (t = -1.66, p = 0.11, \eta^2 = 0.09) \).

Written production posttest. On the written production test, the 18 test items targeting give verbs were evenly divided into the three item types: Self \( (n = 6) \), Family \( (n = 6) \), and In/Out-group \( (n = 6) \). Each type was examined separately to determine whether the Planning Group and the No Planning Group differed in their test scores. The descriptive statistics are presented in Table 16.

Table 16. Written production posttest scores for test-item types by group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Self</th>
<th></th>
<th>Family</th>
<th></th>
<th>In/Out-group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td>Planning</td>
<td>3.71</td>
<td>1.14</td>
<td>3.21</td>
<td>0.80</td>
<td>3.21</td>
<td>0.80</td>
</tr>
<tr>
<td>No Planning</td>
<td>4.38</td>
<td>1.50</td>
<td>3.38</td>
<td>1.26</td>
<td>3.38</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Independent \( t \)-tests revealed no significant differences between the groups on Self items \( (t = -1.34, p = 0.19, \eta^2 = 0.06) \), Family items \( (t = -0.41, p = 0.69, \eta^2 = 0.01) \), or In/Out-group items \( (t = -0.47, p = 0.64, \eta^2 = 0.01) \). A repeated-measures ANOVA with planning condition as the between-subjects factor and item types as the within-subjects factor was performed to determine whether there was an overall difference (regardless of planning group conditions) between the item types, and whether the interaction between the planning group condition and the item type was significant. The analysis revealed that the difference in test scores according to the item types was significant \( (F = 9.50, p < 0.01, \eta^2 = 0.25) \). The interaction was not significant \( (F = 1.06, p = 0.36, \eta^2 = 0.04) \). Post hoc
analysis reported a significant difference between Self and Family ($F = 12.54, p = 0.001, \eta^2 = 0.31$) and also Self and In/Out-Group ($F = 9.90, p = 0.004, \eta^2 = 0.26$). No significance was found between Family and In/Out-Group ($F = 0.00, p = 1.00, \eta^2 = 0.00$). These *post hoc* analysis results indicate that on the written production posttest the participants were significantly more accurate on Self items than on the other two types of items.

The results of the three item types were further tested for each verb. The descriptive statistics are shown in Table 17. Because the number of items for each of the three item types was six and this was further divided by the two verbs, for each cell, the maximum possible score was three.

Table 17. Written production posttest scores for test-item types by verb and group

| Verbs & Groups | Self | | | Family | | | | In/Out-group | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ |
| **Ageru** | | | | | | | | | | | | |
| Planning | 3.00 | 0.00 | 3.00 | 0.00 | 3.00 | 0.00 |
| No planning | 2.75 | 0.58 | 2.88 | 0.50 | 2.81 | 0.75 |
| **Kureru** | | | | | | | | | | | | |
| Planning | 0.71 | 1.14 | 0.21 | 0.80 | 0.21 | 0.80 |
| No planning | 1.63 | 1.41 | 0.50 | 1.10 | 0.56 | 1.21 |

16 The participants in the Planning Group always used *ageru* on the *ageru* items.
According to independent samples $t$-tests, no significant differences were found among the comparisons between the Planning Group and the No Planning Group, whether for the *ageru* items focusing on Self ($t = 1.62, p = 0.12, \eta^2 = 0.09$), Family ($t = 0.93, p = 0.36, \eta^2 = 0.03$), or In/Out-group ($t = 0.93, p = 0.36, \eta^2 = 0.03$), or for the *kureru* items focusing on Self ($t = -1.93, p = 0.06, \eta^2 = 0.12$), Family ($t = -0.81, p = 0.43, \eta^2 = 0.02$), or In/Out-group ($t = -0.915, p = 0.37, \eta^2 = 0.03$).

**Written recognition posttest.** On the written recognition test, the 18 test items targeting *give* verbs were evenly divided into the three item types: Self ($n = 6$), Family ($n = 6$), and In/Out-group ($n = 6$). Each item type was examined separately to determine whether the Planning Group and the No Planning Group differed in their test scores. The descriptive statistics are presented in Table 18.

**Table 18. Written recognition posttest scores of test-item types by group**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Self</th>
<th></th>
<th>Family</th>
<th></th>
<th>In/Out-group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Planning</td>
<td>3.93</td>
<td>1.33</td>
<td>3.43</td>
<td>1.09</td>
<td>3.36</td>
<td>1.15</td>
</tr>
<tr>
<td>No Planning</td>
<td>4.88</td>
<td>1.36</td>
<td>3.63</td>
<td>1.46</td>
<td>3.38</td>
<td>1.41</td>
</tr>
</tbody>
</table>

A series of independent $t$-tests revealed no significant difference for Self ($t = -1.92, p = 0.07, \eta^2 = 0.12$), Family ($t = -0.41, p = 0.68, \eta^2 = 0.01$), or In/Out-group ($t = -0.04, p = 0.97, \eta^2 = 0.00$). In order to determine if there was an overall difference (regardless of planning group conditions) between the item types, and whether there was the interaction between the planning group condition and the item type; a repeated-measures ANOVA.
with planning condition as the between-subjects factor and item types as the within-subjects factor was performed. This analysis showed the difference in test scores according to the item types was significant ($F = 14.19, p < 0.01, \eta^2 = 0.34$) and the interaction was not significant ($F = 2.77, p = 0.07, \eta^2 = 0.09$). A significant difference was found in the *post hoc* analysis between Self and Family ($F = 12.54, p = 0.001, \eta^2 = 0.31$) and also Self and In/Out-Group ($F = 17.44, p < 0.01, \eta^2 = 0.38$). No significance was found between Family and In/Out-Group ($F = 1.55, p = 0.22, \eta^2 = 0.05$). These *post hoc* analysis results indicate that on the written production posttest participants were significantly more accurate on Self items than on the other two types of items, as found in the other two tests.

The results on the three item types were further tested for each verb. Because the number of items for each of the three item types was six and each was divided evenly between the two verbs, for each cell the maximum possible score was three. The descriptive statistics are shown in Table 19.
Table 19. Written recognition posttest scores of test-item types by verb and group

<table>
<thead>
<tr>
<th>Verbs &amp; Groups</th>
<th>Self</th>
<th>Family</th>
<th>In/Out-group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Ageru Planning</td>
<td>3.00</td>
<td>0.00</td>
<td>2.86</td>
</tr>
<tr>
<td>No Planning</td>
<td>2.81</td>
<td>0.75</td>
<td>2.81</td>
</tr>
<tr>
<td>Kureru Planning</td>
<td>0.93</td>
<td>1.33</td>
<td>0.57</td>
</tr>
<tr>
<td>No Planning</td>
<td>2.06</td>
<td>1.29</td>
<td>0.81</td>
</tr>
</tbody>
</table>

According to a series of independent samples $t$-tests, for the *ageru* items, significant differences were not found between the Planning Group and the No Planning Group: Self ($t = 0.93, p = 0.36, \eta^2 = 0.03$), Family ($t = 0.23, p = 0.82, \eta^2 = 0.002$), or In/Out-group ($t = 0.96, p = 0.34, \eta^2 = 0.03$). For the *kureru* items, however, there was a significant difference for Self ($t = -2.37, p = 0.03, \eta^2 = 0.15$), but not for Family ($t = -0.55, p = 0.59, \eta^2 = 0.01$) or for In/Out-group ($t = -0.65, p = 0.52, \eta^2 = 0.01$).

*Summary of results from the three test-item types.* The results from the statistical analyses did not show any significant differences except with the following two comparisons. First, for the In/Out-group items on the oral posttest, the No Planning Group had significantly higher scores than the Planning Group. The effect size was large. Second, for the Self items targeting *kureru* on the recognition test, the No Planning
Group had significantly higher test scores compared to the Planning Group. The $\eta^2$ value indicated a medium effect.

In terms of the overall difference (regardless of planning group conditions) among the item types, for all three tests the participants performed statistically significantly more accurate on Self items than on the other two types of items, Family and In/Out-Group. The $\eta^2$ values indicated a large effect.

Research Question 2:

Do Learners Who Receive Recasts Following Strategic Planning Show Greater Noticing of L2 Linguistic Forms and Pragmatic Factors Than Learners Who Have No Planning Time Available?

In order to answer the second research question, only the data from the strategic planning and stimulated recall group (+P/SR) and the no strategic planning and stimulated recall group (-P/SR) were used. The strategic planning and posttest group (+P/T) and the no strategic planning and posttest group (-P/T) were not included here because they did not participate in the stimulated recall interviews which were conducted to gauge participants’ noticing during the treatment task interactions. For the reporting of the results for Research Question 2, the strategic planning and stimulated recall group (+P/SR) and the no strategic planning and stimulated recall group (-P/SR) will be called the Planning Group and the No Planning Group, respectively.

The four data sources—namely, 1) the verbal reports from the think-aloud protocols during planning, 2) the verbal reports from the stimulated recall interview, 3) the interactions during the two treatment tasks, and 4) the reports from the
questionnaire—were coded for evidence of noticing of linguistic form and noticing of pragmatic concepts related to the two *give* verbs. The coded data revealed that all of the participants in both groups noticed the linguistic form, *kureru*, a form with which they were unfamiliar before this experiment.

Regarding the participants’ noticing of contextual factors involved in the pragmatic principle which guides usage of the two *give* verbs, three facets were coded in this investigation: two relevant contextualized factors (i.e., Self & Non-self and Family & Non-family) and the pragmatic principle itself (i.e., In-group & Out-group). The noticing of these two factors and the pragmatic principle among the participants in the Planning Group and the No Planning Group is summarized in Table 20.

Table 20. Number of participants who noticed contextual factors/pragmatic principle by types and group

<table>
<thead>
<tr>
<th>Groups</th>
<th>Contextual factors and pragmatic principle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self</td>
</tr>
<tr>
<td>Planning (N = 10)</td>
<td>4</td>
</tr>
<tr>
<td>No Planning (N = 10)</td>
<td>4</td>
</tr>
</tbody>
</table>

It is obvious from Table 20 that no difference in noticing was found between the two groups. Thus, a statistical analysis was not performed. One interesting pattern emerging from the results shown in Table 20 is that participants in both groups seemed to be most likely to notice Self & Non-self contextual factor, followed by Family & Non-family, and finally, the more general pragmatic principle. It does not make sense, however, to test for
statistical significance here because the numbers are too small. More detailed data regarding noticing are found in Appendix R in which the noticing data were broken down into the four data sources.

In Discussion Chapter, The quantitative results of the two research questions addressed in this chapter will be explained with qualitative data from think-aloud protocols, the stimulated recall interview, and the questionnaire.
Chapter 4: Discussion

This current experiment addressed two questions:

Research Question 1

Do learners who receive recasts following strategic planning show greater L2 development than learners who have no planning time available?

Research Question 2

Do learners who receive recasts following strategic planning show greater noticing of L2 linguistic forms and pragmatic factors than learners who have no planning time available?

In this chapter, first, the results from Research Question 1 are discussed, and then those of Research Question 2 will be addressed. For Research Question 1, the prediction from the two models of attention and the results from this current experiment will be presented. The results are then discussed with 1) the comments from the stimulated recall interviews and exit questionnaires and 2) other data from the current study, such as participants’ perceptions of task difficulty ratings. The findings from previous strategic planning and task studies will be also used to support my arguments. For Research Question 2, first, the current study’s results regarding noticing are restated, and then the results are explained, noting the possible limitations of noticing measurements this study employed. Based on the noticing results, the efficacy of recast intervention in increasing the chances of learners noticing L2 pragmatic functions is also discussed.
Research Question 1

Prediction from the Two Models of Attention

As mentioned earlier in Chapter Two, Review of Literature, both the single-resource model (Skehan, 1998; Skehan & Foster, 2001) and the multiple-resource model (e.g., Robinson, 2007, 2010) imply that when learners are not given planning time prior to task performance, the amount of attention and memory resources learners have available for focusing on language during the task should be reduced. A logical assumption based on these theoretical arguments is that because the provision of time for strategic planning can increase the availability of attentional and memory resources during task performance, the occurrence of learners’ noticing should increase, facilitating L2 development.

Results from this Study

The results from the current experiment, however, did not appear to follow the outcome predicted using the two attentional models. First of all, significantly more improvement was found in the No Planning Group compared to the Planning Group in the oral posttest. Moreover, regarding the newly introduced word, kureru, the No Planning Group scored higher on kureru items compared to the Planning Group in all posttests, and the results on the oral test showed the difference approached statistical significance. In addition to those test scores, the results on the frequencies of ageru and kureru usage on the posttests also appeared not to provide support for the models’ prediction. While the participants in both the Planning Group and the No Planning Group had strong inclinations toward using the known word, ageru, compared to the newly
introduced word, *kureru*; this pattern was more prominent among participants in the Planning Group than participants in the No Planning Group.

*Discussions of the Results*

So, why did the results regarding L2 development show results rather opposite of what Robinson and Skehan’s theories would predict, and why was the newly introduced word, *kureru*, utilized less often on the posttests among participants in the Planning Group? One possible explanation for these apparently perplexing findings may be that the availability of strategic planning time did not alleviate the burden on participants’ attentional and memory resources when they performed the current study’s treatment tasks. In other words, strategic planning did not create conditions that supported participants’ L2 development in this current experimental design. In order to gain insight into and to support for this argument, I reviewed the stimulated recall data in which participants commented on their thought processes during task performance as well as the comments made by the participants in the Planning Groups (+P/T and +P/SR) in response to the question in the exit questionnaire which asked whether the strategic planning was helpful.

*Comments from stimulated recalls and exit questionnaires.* The participants’ comments related to strategic planning indicated three patterns: 1) a perceived benefit (or benefits) from strategic planning, 2) difficulty in making use of plans made during the task performance due to forgetting what they planned, and 3) additional attentional and memory burdens due to strategic planning. Interestingly none of the participants
mentioned anything related effect of strategic planning on language itself (i.e., grammar, give verbs) in either the questionnaire or stimulated recalls.

The comments regarding the benefits of strategic planning, as shown in Example 15, indicate that the provision of time for strategic planning (reportedly) enabled participants to make better choices at least during the planning and to rehearse their responses.

Example 15. Comments regarding the benefits of strategic planning

Comments from exit questionnaires:

a. P16: *It gave me more time to think about the prices of the different gifts.*

b. P20: *It was helpful to try and guess the prices.*

c. P8: *Looking over my written planned responses [during planning time] helped me compare my answers overall and pick better options.*

d. P3: *I could organize my thoughts and keep track of what I had done, as well as rehearse responses.*

Seven out of 24 participants in the Planning Groups (+P/T and +P/SR) reported in the exit questionnaires benefits of strategic planning similar to the ones in the above example. In stimulated recalls, there were no comments related to the benefits of planning. From the participants’ comments, it is clear that strategic planning provided opportunities for at least some of the participants to prepare for their upcoming task performance.

While there were some reported benefits of strategic planning, some participants’ comments also revealed that they felt that prior preparation for task performance did not
always impact actual task performance because they could not remember what they had planned. Some of these comments are listed in Example 16.

Example 16. Comments suggesting difficulty in making use of what was planned during task performance due to forgetting

Comments from exit questionnaires:

a. P22: *It definitely made it a little easier than the activity without planning [oral pretest], but it wasn’t as easy as I thought because I couldn’t remember the answers I planned.*

b. P1: *It helped a little bit, but I think that there were so many questions, I couldn’t remember the answers I gave the first time.*

c. P9: *It was helpful, but I forgot a lot when the time came anyway.*

Five out of 24 participants indicated the comments similar to the above in exit questionnaires, however no comment related to difficulty in utilizing plans was found in stimulated recalls. The participants’ comments stated that it was extremely difficult to hold everything they had planned in memory and to use that planned information during task performance.

In contrast to Robinson and Skehan’s predictions, the most intriguing comments revealed a perception that suggests strategic planning actually placed more pressure on participants’ attentional and memory resources. More specifically, planning may have led to resource dispersion because the participants were working hard to remember what they planned and/or were worried about remembering their plans. These, in turn, required
extra thinking during task performance and distracted from attending the target forms, give verbs. In Example 17 shows those comments.

Example 17. Comments suggesting additional attentional and memory burden

Comments from exit questionnaires:

a. P21: Slightly helpful since I used the time to go through and have my answers figured out in a way that made sense so I could follow well, but it was hard to remember quickly.

b. P12: While it made it easier to try to figure out the prices relative to the other items, it was more difficult to remember my decisions while recalling.

Comments from stimulated recalls:

c. P21: Oh I was trying to remember what I decided on for the value, so I wouldn’t end up getting messed up toward the end, we got all the different prices and presents …… they are probably the same thing, just trying to remember how it xxx [inaudible], because if I messed up the prices probably lose the continuity in my thought process before so.

d. P22: Um, I was trying to remember what I had done during the planning …. um..I think I was confused about the prices, I couldn’t remember what I wrote... I was forgetting my plan, …. I was panicking definitely, I forgot which I thought and I couldn’t chose, panicking kind of.

I was thinking that I forgot the choice that I was planning to say and about the verb confusion which one should I use.
e. P23: *I was trying to figure out like kuremasu and agemasu, at the same time I was trying to remember what I answered before like what I planned.*

Four out of 24 participants in the Planning Groups made the above comments in exit questionnaires and/or stimulated recalls. The comments from Participant 22 (P22) and Participant (P23) directly indicated they were paying attention to the give verbs while also trying to remember what he/she had planned, showing strategic planning created distraction more than it helped the participants to attend to language. In a previous task study, which examined the provision of contextual support in relation to the benefits of recasts, (Révész, 2009) also argued that the provision of photos (or contextual support) which were claimed to free up learners’ attentional resources may have distracted learners’ attention from the linguistic part of the task, and drawn learners’ attention from the recasts. This current study and Révész’s study suggest that what is believed to free up attentional resources may not always function as expected and may actually deplete the resources.

The comment from Participant 22 (P22) also depicted that the participant was worried about remembering the information she planned. In other words, planning induced anxiety occurred. Similar findings were reported in a study by Ortega (1999, 2005). In that study a participant noted being “a little bit more tense to do it, because getting to write notes [during planning] ‘oh I have to remember more things’” (p. 90). Another participant stated, “When you prepare you get worried. You get nervous that you are going to mess up and then you probably do. That’s what I think” (Ortega, 2005, p. 90). Anxiety has been argued to affect L2 development (Sheen, 2008). Sheen’s recast study
demonstrated that the participants with low-anxiety outperformed the participants with high-anxiety on both immediate and delayed posttests regarding the acquisition of English articles. Based on the comment from this current study’s participant (P22) and the findings from the previous studies, it can be argued that providing strategic planning time may have induced some anxiety among some participants and possibly impeded the participants’ ability to attend to form.

In order to probe this issue deeper, I have looked into the data from the Planning and Posttest Group (+P/T) to examine whether there is any pattern between participants who reported benefits of strategic planning and those who did not in their posttest performances. In the Planning and Posttest Group (+P/T) 1) two participant reported benefit, 2) four reported difficulty in making use of plans, and 3) one reported additional attentional and memory burdens. They all performed at 50 percent accuracy on their pretests (oral, written production, and written recognition pretests). Their posttest accuracy results are found in Table 21.
Table 21. Posttest accuracy and participant’s reports

<table>
<thead>
<tr>
<th></th>
<th>Oral posttest</th>
<th>Written production posttest</th>
<th>Written recognition posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Participant 8</td>
<td><strong>75%</strong></td>
<td><strong>61.11%</strong></td>
<td><strong>66.67%</strong></td>
</tr>
<tr>
<td><strong>Difficulty in making use of plan</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 1</td>
<td><strong>58.33%</strong></td>
<td><strong>61.11%</strong></td>
<td><strong>61.11%</strong></td>
</tr>
<tr>
<td>Participant 3</td>
<td>33.33%</td>
<td><strong>61.11%</strong></td>
<td><strong>61.11%</strong></td>
</tr>
<tr>
<td>Participant 9</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Participant 13</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Attentional and memory burden</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant 12</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Because of the small number of participants addressed in the table above, it is not possible to make a conclusive argument. However, one of the two participants who reported planning to be beneficial showed improvement on her performance across all the tests, only 1 out of 4 participants who reported difficulty in making use of planning demonstrated improvement across all tests, and the participant who reported attentional and memory burdens did not show improvement. In addition, the highest accuracy was found in the participant who reported planning was helpful. These data, therefore, appear
to suggest that when learners perceived the strategic planning to be beneficial, there is a better possibility that planning will support learning.

*Participants’ perception of task difficulties.* In order to shed light on this current study’s rather contradictory results from the attentional models’ predication, the participants’ perceptions of task difficulty was also examined. On the questionnaire, the participants were instructed to rate the treatment task difficulty on a 5-point Likert scale (1-very easy to 5-very difficult). The descriptive statistics for the ratings can be found in Table 22. The mean rating from the No Planning Groups (-P/T and -P/SR) was very slightly higher than that from the Planning Groups (+P/T and +P/SR), indicating that participants in the No Planning Groups perceived the tasks as slightly more difficult. This difference between the groups did not reach significance (*t* = -0.06, *p* = 0.95).

Table 22. Ratings of perceived task difficulty

<table>
<thead>
<tr>
<th>Group</th>
<th><em>M</em></th>
<th><em>SD</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning (<em>N</em> = 24)</td>
<td>2.83</td>
<td>0.82</td>
</tr>
<tr>
<td>No Planning (<em>N</em> = 26)</td>
<td>2.85</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Regarding learners’ perceptions of task difficulty, one of previous studies has produced findings similar to the current study’s result. Wigglesworth (2001) revealed that when learners engaged in unfamiliar or unstructured tasks which involved in giving instructions and negotiations, they reported the tasks were more difficult with planning. However, when engaged in familiar or structured tasks involved in giving instructions and negotiations, they perceived those tasks to be easier with planning than without
planning. Other studies such as by Elder and Iwashita (2005), which used story retelling tasks, did not find a significant difference in learners’ perceptions of task difficulty according to the availability of planning time; however, the planning was perceived to make the task slightly easier. Tavakoli and Skehan (2005), which also used story retelling tasks, found that planners perceived tasks to be significantly easier compared to non-planners.

Considering the results from the participants’ comments from exit questionnaire and stimulated recalls and ratings in the current study in conjunction with findings from previous studies, it seems simplistic to claim that when learners engage in strategic planning, the amount of attention and memory resources they have available for language will necessarily be increased, thereby positively influencing L2 development. I would argue, as Robinson (2001, 2007) has also mentioned, that many other factors affect the link between strategic task planning and the availability of attention and memory resources during task performance. For the particular circumstances in this study, the Planning conditions possibly involved dual-task demands; in other words, the participants were performing the matching task while also trying to remember what they had planned. This factor, +/- single demand, Robinson has claimed, is a resource-dispersing dimension which can result in negative consequences on task performance when complexity is increased along this dimension.

My results do not challenge the theoretical arguments per se, but rather, they suggest that there are many factors to consider before we can link strategic planning to positive impacts on SLA. For example, planning for tasks such as the matching task used
in the current study may have unexpectedly negative effects on L2 production and/or development because they require learners to hold a large amount of non-sequential information in memory during task performance. For tasks such as story retelling, on the other hand, strategic planning may be more likely to function as the theoretical argument claims. In fact, the previous studies (Elder & Iwashita, 2005; Tavakoli & Skehan, 2005) which used story retelling tasks and examined participants’ perception of task difficulties, found that planners perceived task performance to be easier than non-planners. Because story retelling activities involve a logical story line, or linked events, and because learners are given pictures to describe during task performance, the coherence and contextual support may aid learners in keeping planned information in memory and help them to recall it later. In other words, story retelling tasks may involve less demand on remembering the information for later recall. As support for this argument, based on some previous studies of learner production with strategic planning, Foster and Skehan (1996) and Skehan and Foster (1997) speculated that strategic planning may be less beneficial as far as linguistic accuracy is concerned when tasks do not have clear microstructures.

Previous studies have investigated the effects of strategic planning on the performance of different types of tasks (e.g., Foster & Skehan, 1996; Skehan & Foster, 1997). Specifically they focused on the cognitive load of the tasks (e.g., difficult or easy tasks) and its impact on task performance. From the current study’s results, it seems reasonable to argue that researchers must examine tasks from a different perspective, too. By examining learners’ thought processes during task planning and during task
performance, we may discover how different types of tasks shape the effects of strategic planning as far as demands on learners’ attentional and memory resources are concerned. These types of experiments will offer researchers clearer pictures of the effects of strategic planning on learners’ L2 production and development.

Content of Planning and L2 Development

Researchers have argued that there is a need for more studies which examine what learners do during planning (e.g., Batstone, 2005; Ellis, 2009; Mochizuki & Ortega, 2008; Ortega, 1999, 2005; Sangarun, 2005; Skehan & Foster, 2005) and how that which learners planned affects their task performances (Sangarun, 2005). Because this current study collected think-aloud protocols during strategic planning, I decided to present the data from the protocol to provide insight into the above inquiries. The data presented below focus on the protocols gathered during the second strategic planning session which was implemented immediately after the Task One interaction in which participants received recasts on their give verb usage (where they were exposed to the new word, kureru), and before the Task Two performance. The participants’ protocols revealed several patterns in the manners in which participants used the planning time: 1) making choices for the matching task, 2) addressing ageru and kureru, 3) rehearsing their answers using ageru and kureru, and 4) thinking about the possible rules of give verbs. Examples for each are below.
Example 18. Content of planning addressed in think-aloud protocols

Making choices for the matching task

a. P4: Let’s see 20 dollar gift.. um…..salmon, coat, bag, and knife are
towards the 100 + dollars. Sandwich, flowers, ramen, and
daruma is less than 100..so..20dollar gift would most likely be.. probably
daruma? An then, a pack of cigarettes in Tokyo ..probably 4 dollars.. 40
dollar gift, oh, 40 dollars, I would give back , mmm.... probably...ah. knife?

Addressing ageru and kureru

a. P3: Horrible, last one was horrible...agemasu... kuremasu. I have to
remember these for next time because I did horrible

Rehearsing answers with ageru and kureru

a. P5: 映画のチケット、私は森さんに。。。。。20ドルをあげます。
20ドルをあげます。60 dollars, I don’t know. Ah...大森さんは私に
電話をくれます。

Movie ticket, I to Mori..... 20 dollars agemasu. 20 dollars agemasu. 60
dollars, I don’t know. Ah... Omori to me phone kuremasu.

Thinking about the possible rules of give verbs

a. P14: Kuremasu and agemasu, cause kure has to be money, I already know
that, when we are doing with money, kuremasu; don’t know when it
applies to things, Don’t think it applies to things at all, just going with
kuremasu yeah, kuremau, that’s for anything with money, agemasu for
things, that’s an official rule, making it apply now [incorrect rule]
b. P1: So if I’m giving... seems like if I give them something, I’m supposed to say, Kure, Kuremasu. No wait, if someone give something to me, I think it’s kuremasu... but if I give something to someone, agemasu? Maybe..... I think kuremasu. [correct rule]

It is important to mention that the two examples listed for Thinking about possible rules of *give* verbs can also be classified into two patterns, correct rule and incorrect rule.

In Table 23, participants from the Planning and Posttest Group (+P/T) were categorized into four groups based on their improvement or lack of thereof from pretests to posttests. More specifically they were separated in the following three ways: 1) improvement in all tests (oral, written production, and written recognition), 2) improvement on some of the tests, and 3) no improvement. What participants did during task planning is also noted in the table with a checking mark.
Table 23. Test improvement and content of planning

<table>
<thead>
<tr>
<th>Participant</th>
<th>Making choices</th>
<th>Addressing <em>ageru</em> &amp; <em>kureru</em></th>
<th>Rehearsing using <em>ageru</em> &amp; <em>kureru</em></th>
<th>Thinking about rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvement in all tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>√</td>
<td></td>
<td></td>
<td>√ (correct)</td>
</tr>
<tr>
<td>Improvement in some of tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No improvement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>√</td>
<td></td>
<td></td>
<td>√ (incorrect)</td>
</tr>
<tr>
<td>P14</td>
<td>√</td>
<td></td>
<td></td>
<td>√ (incorrect)</td>
</tr>
<tr>
<td>P12</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A few clear patterns emerge from the data in the above table. First, all the participants used the planning time to make choices for the upcoming matching task. In fact, when examining each participant’s protocol, including the participants who also used the time for other purposes, the main focus during the planning was to make choices and be ready for the upcoming task. This finding was not surprising because Sangarun’s (2005) study, which also gathered data during planning using think-aloud protocols, indicated that regardless of the type of task instructions (meaning-focused, form-focused, meaning- plus...
form-focused), the participants showed a clear tendency to attend to content of tasks rather than language itself.

Second, all the participants who demonstrated improvement in all the tests also used the time to focus on the target form, *ageru* and *kureru*, which was shown through their comments that they either rehearsed their answers using *ageru* and *kureru* or thought about the rules. This apparently suggests that attending to form may have produced a positive effect on L2 development. However, the data also indicated some of the participants who did not demonstrated development also spent the time focusing on form. One of the reasons for the paradoxical findings may be that as the data indicated, among the latter participants, when they spent time on talking about rules, their rules were incorrect; therefore it did not aid their development. Another reason may be, as Ellis (2005) predicated, based on the Levelt’s model of speech production, strategic planning may assist conceptualization which results in greater message complexity and better fluency, but not greater accuracy because it does not provide time for the controlled processing required for monitoring. In other words, spending strategic time on focusing on form may not always lead to more attention to form, accurate production, during task production.

**Research Question 2**

**Results from this Study**

Research Question 2 asked if strategic planning has an effect on noticing of L2 forms and contextual factors relevant to pragmatics. All of the participants noticed the verb *kureru*, the new linguistic form related to the targeted pragmatic rule. Because
kureru is a meaning-bearing lexical item located in sentence-final position and participants were corrected whenever they usedageru for kureru items during the two treatment task interactions, there is no doubt that it was quite salient and easy to notice, regardless of planning condition. The contextual factors were, however, not quite as salient; therefore, participants’ noticing was expected to be more susceptible to the planning conditions. Based on the posttests result which showed that 1) the No Planning Group consistently used kureru more frequently than the Planning Group did on the posttests, and 2) the No Planning Group consistently scored higher on the posttests and significantly higher on the oral test, it was expected that more noticing would be found among the No Planning Group. However, no difference was found between the groups in their noticing of contextual factors related to give verbs. Four out of 10 participants in each experimental group noticed Self and/or Non-self contextual factors. Two out of 10 participants in each experimental group noticed Family & Non-family contextual factors. Only one in 10 participants in each group noticed the pragmatic principle. The perplexing finding is that the number of participants who noticed each of the factors was identical across groups; the availability of time for strategic planning appeared not to affect the noticing results at all.

Possible Limitations of Noticing Measurements

One possible explanation for this result may be that the current study’s noticing measurements could not adequately capture the participants’ noticing of the relevant contextual factors. In this study, four measurements were used to gauge participants’ noticing: 1) think-aloud protocols during strategic planning, 2) stimulated recall
interviews focusing on the treatment task interactions, 3) comments made by the participants during the treatment task interactions, and 4) post-experiment questionnaires. Most of the participants’ reports of their noticing of contextual factors came from the stimulated recalls and post-experiment questionnaires; only a few reports of noticing were found in the think-alouds during planning and comments during the treatment task interactions. It was expected that richer data from online measures would be collected compare to offline measures, but this was not the case for this current study. This probably occurred because during the planning time, when the participants were producing their think-aloud protocols, they did not receive feedback, which may have meant that there were fewer opportunities for becoming notice. Moreover, the time provided for planning was limited to 10 minutes, which may have forced participants to focus solely on preparing for the upcoming task performance, leaving little time for them to think about form. Regarding the participants’ comments during the treatment task interactions, it is quite unnatural to imagine that they would readily report their thought processes while performing meaning-focused task activities with a native speaker. The data from the interactions, however, provided important data regarding participants’ noticing of the form, kureru.

As discussed earlier in Chapter Two, Review of Literature, questionnaires and stimulated recall interviews (the sources of most of the reports of noticing in this study) are offline measures and therefore susceptible to memory deterioration. Researchers have

---

17 This may also be a reason that none of the participants mentioned anything related to linguistics concerning benefits of strategic planning (This was addressed in the discussion of Research Question 1). In fact, only a few participants mentioned form during planning and often they did so only after the planning of the content was completed.
also argued that learners’ reports from those measures may be incomplete, or that learners may report what they think they were thinking, instead of the thoughts they actually had at the time of original task performance (e.g., Leow, 2002). In the case of stimulated recall protocols, concerns about memory decay can be reduced with the use of video- and audio- recordings as stimuli for recall (e.g., Bloom, 1954; Gass & Mackey, 2000); however, there are still several implementation factors which might influence learners’ reports. In the current study, in order to make the participants feel more comfortable vocalizing their true thoughts freely, a research assistant conducted the stimulated recall interviews. In order to reduce fabrication of thoughts, she was trained to refrain from asking for reasons or explanations during the interviews. These precautions, however, may have restricted the ability to collect accurate data on participants’ noticing of contextual factors. For example, some participants made comments in the stimulated recall sessions which suggested that they noticed something (see Example 19); however, because there were no follow-up questions on those comments, it was not possible to determine whether noticing of contextual factors/pragmatic principle related to give verbs had occurred. It is quite plausible that some data on participants’ noticing may have been lost due to the precautions taken in the stimulated recall interviews.

Example 19. Comments in stimulated recall

Participants in the Planning Group:

a. P21: *There I was thinking maybe I figured out, and tried to use it, the kuremasu.*

b. P23: *I said kuremasu...just figured out.*
Participants in the No Planning Group:

c. P15: *I think this is the first time I used kureru correctly and after that I knew what it was.*
d. P19: *I kinda figured out when to use kuremasu and agemasu.*
e. P25: *trying to decide like who is kuremasu and agemasu, trying like to figure out the rule and I didn’t get it done until the end*
f. P26: *There were relation to one another to distinguish between kuremasu and agemasu.*

In these quotations, it is evident that the participants noticed something related to the give verbs. However, without further questioning, it was not possible to determine what they noticed.

*Contextual Factors and Pragmatic Principle*

There was no apparent consensus between the development results from the posttests and the noticing results from the stimulated recalls and exit questionnaires: The former suggested No Planning to be a more favorable condition for L2 development, whereas the latter showed no difference between the planning conditions. Nonetheless, there was at least one consistent pattern between the two types of results (however, they were different participants). The noticing results revealed that participants were more likely to notice Self and/or Non-self contextual factors (*n*=4 in each group) than Family & Non-family contextual factors (*n*=2 in each group) and the pragmatic principle (*n*=1 in each group). This pattern was also found in participants’ scores on all three tests, which
showed higher test scores on Self and/or Non-self items over Family & Non-family items and In-group/Out-group items (though no statistical significance was found).

The interesting aspect of the results involving the noticing of relevant contextual factors/pragmatic principle is that the very small number of participants noticed the Family & Non-family contextual factor and especially the general pragmatic principle. These results suggest that even with intensive and focused recasts, noticing of such factor/pragmatic principle was quite limited. It may be that some of factors relevant to pragmatic distinction are not straightforward to learners and solely receiving feedback may not be sufficient for learners to notice the distinction. Schmidt (1993) stated,

[L]earning a new pragmatic system often entails learning how to make new interpretative assessments of the world…. Simple exposure to sociolinguistically appropriate input is unlikely sufficient for second language acquisition of pragmatic knowledge… because the linguistic realizations of pragmatic functions are sometimes opaque to language learners and because the relevant contextual factors to be noticed are likely to be defined differently or may be nonsalient for the learner. Second language learners may fail to experience the crucial noticing for years (pp. 34, 36).

The struggles participants experienced finding out relevant contextual factors/pragmatic principle were often expressed in learners’ stimulated recall. A few of these participants’ comments are provided in Example 20. For each participant, entire comments related to the verbs and looking for some patterns were extracted.
Example 20. Comments about struggling to find patterns

a. P22: Task One: I was like what was kuremasu?[item 3]...I was so confused about the verb, yeah, that’s about it [item4]....I think at this time, I thought when I give it to someone, that should be agemasu, and I was thinking the whole thing about agemasu and kuremasu [item7]...and now I thought when we are talking about money we should use kuremasu [item8]...now I was confused again, we were talking about money but we used the other one [item 9]...I was getting embarrassed, which verb should I chose?[item 10]...I was getting confused what the verbs again [item 11]

Task Two: I was thinking that I forgot the choice that I was planning to say and about the verb confusion which one should I use [item 6]... I was keep thinking about the verbs, trying to guess the rule, and I was thinking maybe something that I have learned, but I don’t know [item 7]

b. P29: Task One: I didn’t know what kuremasu was so I was wondering why it was kuremasu instead of agemasu [item 3]...um again just adding to the confusion of why it’s kuremasu instead of agemasu [item 4]....again confused about the same thing, at this time I was sort of looking for patterns, but [item 5]...Is kuremasu to receive?[item 7]...still didn’t understand why it was kuremasu [item 8]...uh again trying to figure it out [item 10]...um again I think I was just , by that point I was looking for a pattern in the sentence structure that they might help me figure out what, why I’m getting wrong [item 14]
Task Two: I was really looking for the patterns within the texts for some sort of clues to try to get [item 1]…looking for those patterns, so there when I uh had was denwa [phone] so it was an object, I was thinking well it’s maybe relationship between uh giving of an object and kuremasu, and giving money and agemasu [item 2]…so for that one obviously relationship between money and kuremasu [item 5] … um and that again, I was just confused [item 6]…I just started to look for some connection between kuremasu and uh family members? Like if you are giving something to your family member then that’s kuremasu [item 8]…I thought that, so then I was thinking maybe it was the person who actually received it instead of giving it, I don’t know [item 11].

c. P31: Task One: That was the first time I’ve heard kuremasu, so I was trying to figure out what that was [item 3]…I think I was confused, I didn’t know which one to use, so I was still trying to figure out what the difference was [item 5]…I couldn’t figure out the pattern right, I thought it had to do with the price, so I had no ideas what pattern was [item 7]…still price, something about the price [item 8] …I had no idea how the item related to the verbs, I didn’t know it was item’s price or not, just guessed [item 10]

Task Two: I think at that point I basically given up on trying to figure out the pattern [item 2]…I was thinking about like math digit or something, like number of digits, I tried to think syllables not syllables, like sound or something [item 4]…yeah I think I gave up on the number thing, I was trying
to see there was anything else, so I looked at the people, but I couldn’t figure out anything from the people [item 15]

The above comments clearly show that the participants were aware that they were making the errors on the use of *ageru* and *kureru*, and they constantly looked for patterns in order to use them correctly. Participant 31 (P31) tested several hypotheses throughout the treatment but was not able to arrive at noticing of the contextual factors. The results from the current study add some support not only to Schmidt’s claim, but also potentially to argue that some type of explicit instruction may be helpful in promoting pragmatic development for L2 learners. Bardovi-Harlig (2012) noted, the body of research on the effects of instructions on assisting learners in gaining knowledge of L2 pragmatics is still quite small. Therefore, whether instructions are necessary or not is an important area for investigation in future studies.

**Limitations**

In addition to some of the limitations of this current study mentioned in Discussion Chapter, several others must be addressed. First, limitations relate to the tasks and procedures in this study are reported. Second, shortcomings of the overall design of the study are described.

**Tasks and Procedures**

First, more rigorous piloting of the tasks used in this study could have increased the content validity of the tasks. All the tasks were pilot tested with a native speaker and a near native speaker of Japanese in order to verify that their usage of *ageru* and *kureru* matched what was intended in the items in the treatment and tests tasks. Because the
native speaker had taught Japanese for five years in universities and the near native speaker had studied Japanese for four years in college, both knew the pragmatic rules associated with Japanese give verbs. I assumed this would not have caused any critical difference in their judgments of the usage of ageru and kureru compared to other Japanese speakers; however, if I had included native speakers of Japanese who did not know the pragmatic rules for this piloting, I may have been able to further assure that the items used in this study’s tasks accurately reflect how ageru and kureru are used among Japanese speakers.

Second, richer and more exposure to the target form during treatment session may have been necessary to support noticing and L2 development in this current study. In order to limit any external influence (e.g., contents covered in their Japanese classes, and participants looking up information related to the target pragmatic principle/rules outside the experiment) the treatment session and the posttests were conducted on the same day; therefore, it was not possible to provide more time for participants to be exposed to the target forms. In addition, because of the time constraints and in order to facilitate learning of the target form I only focused on limited situations: 1) Self (in-group) and acquaintance (out-group), and 2) family members (in-group) and acquaintance (out-group). These situations, even though necessary, may have restricted participants’ opportunities to notice the pragmatic concepts (in-group and out-group concept associated to give verbs) and to learn the target form.

In addition, because during the treatment session participants were only exposed to the situations: 1) self (in-group) and acquaintance (out-group), and 2) family members
(in-group) and acquaintance (out-group), it was too much of a stretch to assume that participants would be able to learn the in-group and out-group principle based on these limited situations and therefore correctly answer posttests’ generalization items which incorporated a situation between friend (in-group) and celebrity (out-group). Based on conversations with the Japanese teachers from whom the participants were taking a class, the participants had not learned the in-group and out-group concept and probably did not know what would be considered in-group and out-group. Therefore, the exposure to the limited situations during treatment was probably not sufficient for the participants to answer the generalization items correctly.

Third, mentioning and showing a picture of a family on the first day of the experiment may have facilitated participants’ noticing and L2 development. On the first day of the experiment, at the beginning of the pretests, I introduced people appearing on the pretests using the same pictures used in the test and explained that the person was referred to as *watashi* (i.e., I, me, my) and the family members on the test represented the participant and the participant’s family. This step was not included during the second day of the treatment when they received recast treatment; however, it could be argued that this primed some participants and aided noticing of one of the contextual factors, family, during the second day when they received recast treatment. Because the first day of the experiment was conducted at least a week before the second day of the experiment, I assume, however, the priming effect was very minimal if it even occurred at all.

Forth, the difference in the pictures between the family members and non-family members may have inadvertently assisted participants in more accurately completing
posttests. The pictures used for *watashi* (i.e., I, me, my) and the family members were computer drawings; however, the pictures used for acquaintances and celebrities were real pictures of people. I used the computer drawing for the pictures of *watashi* and family members in order to make them less ethnically specific because the participants in this current study were from a variety of ethnic backgrounds. Because *watashi* and family members in this experiment are in-group and the acquaintances and celebrities are out-group, simply noticing the difference in the types of pictures, rather than the in-group and out-group principle, could have assisted participants in more accurately completing the posttests. There were no comments related to this in any of the qualitative data; therefore, I assumed this did not occur, but the pictures need to be more consistent if the current tasks are to be used again.

Fifth, introspective measures should have been implemented during posttests. In this current study, introspective measures (i.e., think-alouds and stimulated recalls) were used to observe participants’ thought processes during strategic planning and task interactions in treatment session; however, no such measures were utilized for pretests and posttests. If introspective measures were used, especially for the posttests, it would have provided important information, such as how participants chose one verb over the other and whether participants noticed any of the contextual factors and/or the pragmatic principle during the tests. Unfortunately this information is missing from this current study.

Sixth, more numbers of items in each test could have provided a more accurate picture of the outcome of the experiment. Again, due to the time constraints of the
experiment, the numbers of items in each test needed to be fewer than ideal. This resulted in a small number of items especially for each of the three test-item types (i.e., self, family, and in-group/out-group).

Seventh, regarding the results from the participants’ perception of task difficulty, it is possible to argue that a significant difference was not found because the results were based on a 5-point Likert scale. More difference between the groups may have been observed if this current study had used 1) an even number scale, such as a 6-point scale, that would force participants to take a position as to whether they thought strategic planning was helpful, or 2) a higher number scale, such as 7-point scale. In addition, Révész, Michel, & Gilabert (2012) found that asking participants about their perceptions of mental effort may be a better predictor of the cognitive load required of an individual to complete his or her task compared to asking his or her perceptions of task difficulty. This current study’s data was collected by asking participants’ perceptions of task difficulty, and therefore, may not reflect the cognitive load required of participants for completing the matching task.

Eighth, it would have been ideal to report reliability statistics for all the pretests and posttests. Unfortunately, the numbers of the item for each test was small and there was little variation in scores (especially for the pretest due to the participant criteria); therefore, it was not possible to calculate an internal consistency reliability coefficient for the tests used in this current study.
**Overall Design of the Study**

First, because it was extremely difficult to find participants who fit into the required criteria, this current study was not able to include control groups (i.e., a group without recasts and a group which only took the pretests and posttests) or an experimental group without think-alouds. I would like to argue that the gain learners demonstrated was due to the provision of recasts or the treatment itself. However, of course, without control groups, it cannot be conclusively claimed that this was indeed the case. Regarding the think-alouds, although the Bowles’ meta-analysis (2010) showed the issue of reactivity of think-alouds in verbal tasks was minimal, it would have been ideal to have a group without think-alouds during strategic planning to demonstrate this current study’s think-alouds were nonreactive. Because such data are not available in this study, some may argue that think-alouds altered the participants’ planning activities and claim that this may have confounded the effects of planning on learners’ noticing and L2 development.

Second, this current study should have examined individual differences, specifically working memory. Because many of the learners addressed having problem remembering what they planned, it could be argued that learners with larger working memory benefitted more from planning. Not only working memory but other factors, which I argued in the Review of Literature Chapter may have affected the efficacy of recasts and how strategic planning supports form-focused performance, must be systematically examined in order to determine whether manipulating those factors would affect learners’ noticing and L2 development.
Third, as with many other previous studies, the number of the participants was small, and the number of participants in each cell to address noticing issue was extremely small. The small numbers may have restricted this study’s ability to accurately show how planning affected learners’ noticing.

Considering the above limitations, this current study’s results can be only suggestive. Future studies must overcome the limitation of this current study in order to obtain more robust results. Also, as mentioned earlier, in future studies the factors argued to affect the efficacy of recasts and the effects of strategic planning to bring learners’ attention to form during task performance should be systematically examined. With such examinations we would be able to more accurately determine whether strategic planning affects the efficacy of recasts on L2 development.

Conclusion

Recent recasts studies have indicated that task variables may affect the extent to which learners benefit from recasts. In order to extend this line of research, this dissertation examined whether strategic planning, a well attested task variable in the literature regarding task-based learning, promotes L2 development and whether it has an effect on noticing of target forms. The results for L2 development showed the No Planning Groups were more accurate compared to the Planning Groups across all immediate posttests, and one of the posttests yielded a statistically significant difference. The results, therefore, suggest that strategic planning did not create conditions that supported participants’ L2 development in this current experimental design. Considering 1) these results, 2) participants’ comments from exit questionnaires and stimulated recalls,
3) ratings of learners’ perceptions of task difficulty in this current study, and 4) findings from previous studies, it seems simplistic to claim that when learners engage in strategic planning, the amount of attention and memory resources available for language will necessarily be increased, thereby positively influencing L2 development--which was predicated from the two attentional models. For example, planning for tasks such as the matching task used in this current study may unexpectedly negatively affect L2 development because these tasks require learners to retain a large amount of non-sequential information in memory during task performance. It appears that there are many factors to consider before we can link strategic planning to positive impact on SLA. This current study’s results, particularly, suggest that examining learners’ thought processes during task planning and task performance may be important in discovering how different types of tasks shape the effects of strategic planning as far as demands on learners’ attentional and memory resources are concerned.

Regarding noticing, no difference was found between the two groups. Close examination of the stimulated recall protocols revealed that this may have been due to the manner in which the stimulated recall interviews were conducted in this study and suggest precautions should be taken in any future studies intending to use these protocols. The data regarding noticing, however, revealed some interesting insights regarding recasts and noticing of factors related to pragmatics. The data indicated that for certain contextual factor, only a very small number of participants noticed, even with intensive and focused recasts. Stimulated recall protocols revealed that many of the participants experienced difficulty identifying relevant contextual factors/pragmatic principle. The
comments from participants clearly showed they were aware of making errors regarding the use of *ageru* and *kureru*, and consistently, they unsuccessfully looked for patterns in order to use them correctly. Based on the protocols, it may be that some factors relevant to pragmatic distinction are not obvious to learners, and solely receiving feedback may not be sufficient for learners to notice the distinction. The results from this current potentially indicate that some type of explicit instruction may be helpful in promoting pragmatic development for L2 learners.
**PROJECT TITLE**  
*Strategic planning, recast, noticing, and development*

**PROJECT DIRECTOR**  
Mika Hama

**PRINCIPAL INVESTIGATOR**  
Mika Hama

**TELEPHONE**  
202-378-3858

The Georgetown University Institutional Review Board (IRB) has approved this research project. For information on your rights as a research subject, call the Institutional Review Board office at 202-687-1506.

**INTRODUCTION**
You are invited to consider participating in a research study to investigate how language learning occurs during conversation. This form will describe the purpose and nature of the research, its possible risks and benefits, other options available to you, and your rights as a participant in the study. Please take whatever time you need to discuss the study with your family and friends, or anyone else you wish to. The decision to participate, or not to participate, is yours. If you decide to participate, please be sure to sign and date the last page of this form.

**WHY IS THIS RESEARCH STUDY BEING DONE?**
In this research study, I am investigating how adults learn a second language through conversations.

**HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?**
About 60 people will take part in this study. Participants in the study are referred to as “subjects.”

**WHAT IS INVOLVED IN THE STUDY?**
First, you will take a brief test of your Japanese ability. The first part (about 20 minutes) will involve interacting with a researcher using a communicative task in which you will be making choices among several items, and the second part (about 20 minutes) will
involve written questions in which you are asked to fill in a blank to make grammatical Japanese sentences.

Then, 1 week after the test, you will meet with a native speaker of Japanese and you will engage in language activities which are similar to the communicative tasks you will do during the test phase. Each activity will take about a half hour. Your conversations will be video-recorded.

Finally, you will take either an interview with the researcher regarding the activities you did with the native speaker of Japanese, or two more tests like the one that you took on the first day. In this way, it will be possible to see whether the activities that you did with the native speaker of Japanese helped you to develop your Japanese skills.

**HOW LONG WILL I BE IN THE STUDY?**
I expect that you will participate in two session of 40-60 minutes over a two-week period.

The investigator may stop the study or take you out of the study at any time she judges it is in your best interest (e.g., if you experience an injury or if you do not comply with the study plan) or for a variety of other reasons. She can do this without your consent.

You can also stop participating at any time and for any reason. However, if you decide to stop participating in the study, I encourage you to talk to the researcher first.

**WHAT ARE THE RISKS OF THE STUDY?**
No known risks are anticipated with the research.

**ARE THERE BENEFITS TO TAKING PART IN THE STUDY?**
It is reasonable to expect the following benefits from this research: (1) that your Japanese abilities will improve. However, I cannot guarantee that you will personally experience benefits from participating in this study. Others may benefit in the future from the information I obtain in this study.

This research study does not provide treatment for any disorder or condition. Should subjects request such treatment, they will be referred to treatment alternatives outside this study.

**WHO CAN PARTICIPATE IN THE STUDY?**
This study is designed for adult learners of Japanese as a foreign language.

**WHAT ABOUT CONFIDENTIALITY?**
Your name will not be used when data from this study are published.

Every effort will be made to keep your research records and other personal information confidential. However, I cannot guarantee absolute confidentiality.
Individuals from the Georgetown University IRB, other Georgetown University offices, and Federal regulatory agencies may look at records related to this study, both to assure quality control and to analyze data. Your name and any material that could identify you will remain confidential except as may be required by law.

I will take the following steps to keep information about you confidential, and to protect it from unauthorized disclosure, tampering, or damage:

If you agree to participate in the study, you will be randomly assigned an identification number. Throughout the study, all of your data (for example, test scores and audio/video-recordings) will be associated only with this randomly assigned identification number. The list of real names and assigned numbers will be kept in a password-protected file on the researchers’ password-protected computers. If the results of this study are published or presented, your name will never be used. You will be referred to either with a number (for example, Participant #25) or with a fake name. I will use the audio/video-recordings only to analyze the data. I will not share them with anyone else and keep them in a password-protected computer which the researcher can only access to.

**WILL I BE PAID FOR PARTICIPATING?**
You will not be paid for participating in this study.

**WHAT ARE MY RIGHTS AS A RESEARCH PARTICIPANT?**
Participation in this study is entirely voluntary at all times. You have the right not to participate at all or to leave the study at any time. Deciding not to participate or choosing to leave the study will not result in any penalty or loss of benefits to which you are entitled, and it will not harm your relationship with Georgetown University or any of its employees.

Should you decide not to be included in the research, you are requested to notify the investigator about the cancellation. You can contact the investigator by e-mail, through a letter, or in person.

Throughout this study, researcher will tell you about new information related to the interventions in the study, interventions that may be appropriate for you, or any other information that may affect your interest in remaining in the study.

**WHOM DO I CONTACT IF I HAVE QUESTIONS OR PROBLEMS?**
Call Mika Hama (202-378-3858) during regular business hours if you have any questions about the study, any problems, unexpected psychological discomforts, or think that something unusual or unexpected is happening.

Call the Georgetown University IRB Office at 202-687-1506 with any questions about your rights as a research participant.
Statement of Person Obtaining Informed Consent

I have fully explained this study to the subject. I have discussed the study’s purpose, its experimental and nonexperimental procedures and interventions, the possible risks and benefits, the standard and research aspects of the study, the alternatives to participation, and the voluntary nature of participation. I have invited the subject to ask questions and have answered any questions that the subject has asked.

________________________________________
Signature of Person Obtaining Informed Consent   Date

Consent of Subject (or Legally Authorized Representative)

I have read the information provided in this Informed Consent Document (or it was read to me by ____________________________).

My questions were answered to my satisfaction.

I voluntarily agree to participate in this study.

________________________________________
Signature of Subject   Date

________________________________________
Signature of Legally Authorized Representative   Date

Where Appropriate

Upon signing, the subject or the legally authorized representative will receive a copy of this form, and the original will become part of the subject’s clinical record. If there is no relevant clinical record, the original will be held in the subject’s research record.
Appendix B
Instructions and trial items for the oral tasks

Oral pretest:

Instructions for oral tasks

PP

You are going to do an activity in which you will be asked to choose appropriate gifts and actions from a list.

Please make sure that you have two sets of documents:
- Your activity sheets
- The list of possible gifts and actions

For each question in the activity, there is only 1 correct answer. At the end of the activity, the researcher will check how many you got correct. If you get everything correct, you’ll get a prize!😊

Go to next slide ➔
Example 1

Here, Tom gave $20 to Hama. In return, she needs to give Tom a gift which is appropriate to that amount of money.

Which gift will Hama give to Tom in return?  
はまさんはトムさん( )______________。

Choose 1 item from the list and complete the sentence above.

[Images of book, television, coffee]
Example 1

Which gift will Hama give to Tom in return?
はまさんはトムさんに__________。

Answer:

Example 2

Here, Tom is going to Hawaii. What will you lend him?

What will you lend him?
トムさん(__________)。
Example 2

What will you lend him?
トムさん[ ]__________。

*Choose 1 item from the list and complete the sentence above.*

Do you have any questions?

If not, let’s begin!
Instructions for oral tasks

TR

You are going to do an activity in which you will be asked to choose appropriate gifts and amounts of money from a list.

Please make sure that you have two sets of documents:
- Your activity sheets
- The list of possible gifts, and amounts of money

For each question in the activity, there is only 1 correct answer. At the end of the activity, the researcher will check how many you got correct. If you get everything correct, you’ll get a prize!😊

Go to next slide ➔
Example 1

Here, Tom gave $20 to Hama. In return, she needs to give Tom a gift which is appropriate to that amount of money.

Which gift will Hama give to Tom in return?
はまさんはトムさんに______________。

Choose 1 item from the list and complete the sentence above.
Example 1

Which gift will Hama give to Tom in return?
はまさんはトムさんには____________。

Answer:

Example 2

Here, Tom gave a book to Hama. In return, she needs to give Tom an amount of money which is appropriate for that gift.

How much will Hama give to Tom in return?
はまさんはトムさんには____________。
Example 2

How much will Hama give to Tom in return?
はまさんはトムさんに____________。

Choose an amount of money from the list and complete the sentence above.

$800  $20  $2

Answer: $20
Do you have any questions?

If not, let’s begin!
Appendix C
List of people who appeared on the oral tasks

父(ちち)  母(はは)
姉(あね) 兄(あに) 私(わたし) 弟(おとうと) 妹(いもうと)
ちぱ  くどう  あべ  なりた
オプラ ウィル 友だち 友だち
Appendix D
Instructions for written tests

Production test

[Instructions]

- For each of the following items, please read the English sentence and fill in the missing particle and verb in Japanese.

- Write your answer on the answer sheet.

- You are NOT allowed to go back to any of the previous items.

Recognition test

[Instructions]

- For each of the following items, please read the English sentence and choose a particle and a word from the word bank to fill in each blank in the Japanese translation provided.

- Mark your answer on the answer sheet.

- You are NOT allowed to go back to any of the previous items.
Appendix E
Think-alouds practice task

Please solve the math problems below. Please speak your thoughts aloud while you work on the math problems. You can use English, Japanese, or both—whatever feels most comfortable at the time.

1. 182×41=

2. 294×32=

3. 596×84=
Appendix F
Treatment task instructions for the No-Planning Groups

You are going to do an oral activity in Japanese with Mika which is similar to what you did in your last meeting with her.

When you do the activity, please think about

1) making appropriate choices of gifts and amounts of money,
2) communicating naturally and fluently, and
3) using language that is grammatically correct
Appendix G
Stimulated recall instructions

You are going to watch task interaction, you just had. We are interested in what you were thinking at the time you were doing the task. We can hear what you saying by watching the video, but we don’t know what you were thinking. So, what we would like you to do is to tell us whatever you were thinking during the interaction, but not what you are thinking now.

We will show you some parts of the interaction. When I pause the video, please articulate in English or Japanese whatever you were thinking at the time in the interaction, not what you are thinking now. There is no right or wrong answer. Your response can be about anything. Your response can be as long as or as short as you want it to be.

If you don’t remember what you were thinking at that time, you can say “I don’t remember”. If you were not thinking anything particularly, you can say “I wasn’t thinking anything”.
Appendix H  
Biodata questionnaire

<table>
<thead>
<tr>
<th>Name ____________________</th>
<th>Age ____________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender: M F</td>
<td>Native language _______________</td>
</tr>
</tbody>
</table>

Lived in Japan? | Yes / No | When? | How long? |
- | - | - | - |

Visited Japan? | Yes / No | When? | How long? |
- | - | - | - |

**Foreign languages**

<table>
<thead>
<tr>
<th>Languages</th>
<th>How long have you studied? (years/months)</th>
<th>At where?</th>
<th>Your proficiency level (beginner, intermediate, advance, native)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>___________________</td>
<td>__________</td>
<td>___________________</td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
<tr>
<td></td>
<td>___________________</td>
<td>__________</td>
<td></td>
</tr>
</tbody>
</table>
Appendix I
Treatment tasks

Task 1

- Which gift will you give to Miyake-san in return?
  みやけさん

- How much will you give to Takahashi-san in return?
  たかはしさん

- Which gift will Endou-san give you in return?
  えんどうさん
How much will Saitou-san give you in return?
さいとうさんは私に_________。

How much will you give to Endou-san in return?
私はえんどうさんに______。

Which gift will Miyake-san give you in return?
みやけさんは私に____________。
Which gift will you give to Inoue-sanh in return?
私はいのえさんに________。

How much will Takahashi-san give you in return?
たかはしさんは私に________。

How much will you give to Saitou-sanh in return?
私はさいとうさんに________。
Which gift will Arai-san give you in return?
あらいさんは私に____________。

Which gift will you give to Hara-san in return?
私ははらさんに____________。

Which gift will you give to Maeda-san in return?
私はまえださんに____________。
How much will Inoue-san give you in return?
いのうえさんは私に______。

How much will you give to Arai-san in return?
私はあらいさんに_________。

Which gift will Maeda-san give you in return?
まえださんは私に_________。
How much will Hara-san give you in return?

はらさんは私に_______。
List of things to give

サンドイッチ  かばん  コート
日本のサーモン  花（はな）  ラーメン
日本のナイフ  だるま (Japanese wish doll)
(8” blade)  (13” tall)

List of money amounts to give

$3  
$17  
$20  
$60  
$100  
$150  
$300  
$700
How much will you give to Mori-san in return?
私はもりさんに______。

Which gift will Oomori-san give you in return?
おおもりさんは私に______。

Which gift will your mother give to Kajino-san in return?
はははかじのさんに______。

えいがのチケット (movie ticket in Tokyo)
$60

$400
Which gift will your father give to Oomori-san in return?
ちちはおおもりさんに__________。

How much will Mori-san give to your older brother in return?
もりさんはあにに__________。

Which gift will Kuroki-san give to your younger brother in return?
くろきさんはおとうとに____________。
How much will your younger sister give to Ono-san in return?
いもうとはおのさんに

How much will Koga-san give you in return?
こがさんは私に

Which gift will you give to Hamada-san in return?
私は はまださんに
How much will your younger brother give to Gotou-san in return? おとうとはごとうさんに——。

Which gift will your older brother give to Kuroki-san in return? あにはくろきさんに——。

How much will Kajino-san give to your mother in return? かじのさんは ははに——。
How much will your older sister give to Koga-san in return?
あねはこがさんに________。

Which gift will Ono-san give to your father in return?
おのさんはちちに____________。

Which gift will Gotou-san give to your older sister in return?
ごとうさんはあねに____________。
How much will Hamada-san give to your younger sister in return?
はまださんはいもうとに____________。
List of things to give

- Telephone (電話)
- Video camera (ビデオカメラ)
- Bonsai (ぼんさい)
- IKEA table (IKEA のつくえ)
- Umbrella (かさ)
- Pikachu (ピカチュウ)
- GW umbrella (GW のかさ)
- Ice cream (アイスクリーム)
- Radio cassette (ラジカセ)

List of money amounts to give

$2
$20
$25
$40
$150
$200
$300
$700
Appendix J
Strategic planning instructions for the Planning Groups

You are going to do an oral activity in Japanese with Mika which is similar to what you did in your last meeting with her. This time you are going have 10 minutes for planning before you do the activity. Please use the time to prepare for the upcoming activity.

Please use the time to think about
1) making appropriate choices of gifts and amounts of money,
2) communicating naturally and fluently, and
3) using language that is grammatically correct

You won’t be able to use any written notes during this activity—sorry! But you are free to take notes now in case it helps you. Please use this note-sheet (not the task sheets) to prepare what you would like to say.

Notes:
(Please speak your thoughts aloud while you plan. You can use English, Japanese, or both—whatever feels most comfortable at the time 😊)
Appendix K
Oral pretest

What will you lend her?
サラさん（ ）__________________________。

What will you lend him?
トムさん（ ）__________________________。

Which gift will you give to Chiba-san in return?
私はちばさん（ ）__________________。

サラ
トム
ちば
私

あした うち
明日 家でパーティーがあります。

かぜをひきました。

明日家でパーティーがあります。

$20
What will you lend her?
ナタリーさん ( ) _________________________________________

What will you lend him?
マイクさん ( ) _________________________________________

Which gift will Kudou-san give to your younger sister in return?
くどうさんはいもうと ( ) ___________________
Which gift will your friend give to Oprah in return?
(私の)友だちはオプラ(  )__________。

What will you lend her?
エイミーさん(  )______________________________________。

Which gift will Will give to your friend in return?
ウィルは(私の)友だち(  )______________。

Which gift will your father give to Abe-san in return?

ちちはあべさん（ ）______________。

What will you lend him?

ケンさん（ ）______________________。

Which gift will Narita-san give you in return?

なりたさんは私（ ）__________________。

雨がふっています。
List of things to give

ハンバーガー  ペン  ぎゅうにゅう

コピーき  時計(とけい)  じてんしゃ

List of things to lend

お金  本（ほん）  車（くるま）

じしょ  セーター  iPod
Appendix L
Written production pretest items

1. Whoopi Goldberg will lend (my) friend a dictionary.
   ウーピー・ゴールドバーグ( )友だちに辞書を______。

2. (My) friend will lend Tom Cruise a pencil.
   友だちはトム・クルーズ( )えんぴつを______。

3. Tanaka-san will give (my) sister a bike.
   田中さんは( )妹に自転車( )______。

4. Ogawa-san will ask (me) (my) birthday.
   小川さんは( )誕生日( )______。

5. Lindsey Lohan will ask (my) friend the meaning of a word.
   リンジー・ローハン( )友だちに言葉の意味を______。

6. (My) brother will ask Tanaka-san (her) name.
   弟は田中さん( )名前を______。

7. (My) friend will give Lindsey Lohan a book.
   友だちはリンジー・ローハンに本( )______。

8. Ito-san will teach (my) father Kanji.
   伊藤さんは父( )漢字を______。

9. Shimizu-san will teach (me) Hiragana.
   清水さんは( )ひらがな( )______。

10. Yamada-san will ask (my) sister (her) address.
    山田さん( )妹に住所を______。

11. (I) will give Ogawa-san tea.
    ( )小川さんにお茶( )______。

213
12. (I) will teach Yoshida-san English.
   (わたしは)吉田さん( )英語を______。

13. (My) brother will give Yamada-san a pen.
   (弟)山田さんにペン( )______。

14. (I) will lend Sasaki-san an umbrella.
   (私は)佐々木さん( )かさを______。

15. (My) sister will lend Sato-san money.
   (姉)佐藤さんにお金を______。

16. Kimura-san will give (me) flowers.
   木村さんは(私に)花( )______。

17. (I) will ask Kimura-san the time.
   (私は)木村さんに時間( )______。

18. Tom Cruise will give (my) friend a present.
   トム・クルーズは友だち( )プレゼントを______。

19. (My) friend will teach Ashton Kutcher English.
   友だち( )アシュトン・カッチャーに英語を______。

20. (My) friend will ask Eva Longoria (her) phone number.
   友だちはエバ・ロンゴリアに電話番号( )______。

   加藤さん( )兄にかばんを______。

22. Hayashi-san will lend (me) a camera.
   林さんは(私に)カメラ( )______。

23. Patrick Dempsey will teach (my) friend politics.
   パトリック・デンプシーは友だち( )政治を______。

214
24. (My) mom will teach Suzuki-san an English word.

母( )鈴木さんに英語の言葉を______。

Answer sheet for written production pretest

1. ( ) ______________________
2. ( ) ______________________
3. ( ) ______________________
4. ( ) ______________________
5. ( ) ______________________
6. ( ) ______________________
7. ( ) ______________________
8. ( ) ______________________
9. ( ) ______________________
10. ( ) ______________________
11. ( ) ______________________
12. ( ) ______________________
13. ( ) ______________________
14. ( ) ______________________
15. ( ) ______________________
16. ( ) ______________________
17. ( ) ______________________
18. ( ) ______________________
19. ( ) ______________________
20. ( ) ______________________
21. ( ) ______________________
22. ( ) ______________________
23. ( ) ______________________
24. ( ) ______________________
Appendix M
Written recognition pretest items

1. Eva Longoria will ask (my) friend the time.
   エバ・ロンゴリアは友だち( )時間を______。

2. (My) mom will give Yoshida-san a letter.
   母は吉田さん( )手紙を______。

3. (My) brother will lend Kato-san a bag.
   兄は加藤さんにかばん( )______。

4. Ito-san will give (me) a cat.
   伊藤さん( )（私に）ねこを______。

5. My friend will lend Whoopi Goldgerg umbrellas.
   友だちはウーピー・ゴールドバーグにかさ( )______。

6. (I) will give Suzuki-san a bike.
   (私は)鈴木さん( )自転車を______。

7. Sasaki-san will lend (me) a dictionary.
   佐々木さんは(私に)辞書( )______。

8. Sato-san will lend (my) sister money.
   佐藤さん( )あねにお金を______。

9. (My) sister will ask Yamada-san the meaning of a word.
   妹は山田さん( )言葉の意味を______。

10. (I) will lend Hayashi-san (my) pencil.
    (私は)林さんにおいんぴつ( )______。

11. Tom Cruise will lend (my) friend a camera.
    トム・クルーズは友だち( )カメラを______。

216
12. Tanaka-san will ask (my) brother (his) birthday.
田中さんは弟( )誕生日を______。

13. (I) will teach Shimizu-san American politics.
(私は)清水さんにアメリカの政治( )________。

14. Shimizu-san will give (my) dad a movie.
清水さん( )父に映画を______。

15. (I) will ask Ogawa-san (his) address.
(私は)小川さん( )住所を______。

鈴木さんは母に日本語( )________。

17. (My) friend will ask Lindsey Lohan (her) name.
友だち( )リンジー・ローハンに名前を______。

18. (My) friend will teach Patrick Dempsey an Japanese word.
友だちはパトリック・デンプシーに日本語の言葉( )______。

19. Kimura-san will ask (me) (my) phone number.
木村さん( )(私に)電話番号を______。

20. Ashton Kutcher will teach (my) friend how to use a camera.
アシュトン・カッチャーは友だちにカメラの使い方( )______。

21. Yoshida-san will teach (me) Kanji.
吉田さんは(私に)漢字( )______。

22. (My) friend will give Ashton Kutcher a TV.
友だち( )アシュトン・カッチャーにテレビを______。

23. (My) dad will teach Ito-san English.
父は伊藤さん( )英語を______。

217
24. Lindsey Lohan will give (my) friend a watch.
リンジー・ローハンは友だち( )時計を______。

Answer sheet for written recognition pretest

<table>
<thead>
<tr>
<th>Word bank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. すすめる</td>
</tr>
</tbody>
</table>

Particle: は に へ を と

1. ( ) _____________________    13. ( ) _____________________
2. ( ) _____________________    14. ( ) _____________________
3. ( ) _____________________    15. ( ) _____________________
4. ( ) _____________________    16. ( ) _____________________
5. ( ) _____________________    17. ( ) _____________________
6. ( ) _____________________    18. ( ) _____________________
7. ( ) _____________________    19. ( ) _____________________
8. ( ) _____________________    20. ( ) _____________________
9. ( ) _____________________    21. ( ) _____________________
10. ( ) _____________________    22. ( ) _____________________
11. ( ) _____________________    23. ( ) _____________________
12. ( ) _____________________    24. ( ) _____________________
Appendix N
Oral posttest

Which gift will the Madonna give to your friend in return?
マドンナは友だちに__________。

How much will Miyagi-san give you in return?
みやぎさんは私に__________。

What will you lend him?
ケンさんに__________________。

友だち
マドンナ

$80

すもうのチケット
(Sumo ticket)

私
みやぎ

ケン

かものいい
買い物に行きます。
Which gift will Sakamoto-san give to your older brother in return?
さかもとさんはあにに__________。

Which gift will your friend give to John in return?
友だちはジョンに__________。

What will you lend her?
エイミーさんに______________________。

ふられました。。(I was dumped...)
How much will you give to Sakamoto-san in return?
わたしさかもとさんに______。

Which gift will your mother give to Miyagi-san in return?
はははみやぎさんに__________。

Which gift will Matsuo-san give you in return?
まつおさんは私に____________。
マイクさんには$70。

Which gift will you give to Yamaguchi-san in return?
私はやまぐちさんに$70。

ナタリーさんに明日BBQパーティーをします。

What will you lend him?
マイクさんにあした図書館で勉強します。
How much will your older brother give to Matsuo-san in return?
あにはまつおさんに__________。

What will you lend her?
サラさんに__________________。

What will you lend him?
トムさんに__________________。

NYへ行きます
コンピューターがこわれました。
How much will Madonna give to your friend in return?
マドンナは友だちに__________。

How much will Yamaguchi-san give to your mother in return?
やまぐちさんはははに_____________。

How much will your friend give to John in return?
友だちはジョンに_____________。
List of things to give

- 時計（とけい）
- テント
- 酒（さけ）
- こい
- いす
- せんたくき

List of money amounts to give

- $20
- $30
- $40
- $100
- $300
- $700

List of things to lend

- (ノート)
- コンピューター
- CD
- クレジットカード
- iPhone
- カメラ
Appendix O
Written production posttest items

1. (I) will teach Shimizu-san English.
( 私 は) 清水さんに英語を______。

2. (I) will give Sasaki-san a watch
( 私 は) 佐々木さんに時計を______。

3. Kato-san will give (me) a pen.
加藤さんは( 私 に) ペンを______。

4. (My) mom will give Suzuki-san water.
母は鈴木さんに水を______。

5. (My) friend will ask Lindsey Lohan (her) birthday.
友だちはリンジー・ローハンに誕生日を______。

6. Tom Cruise will lend (my) friend a pencil.
トム・クルーズは友だちにえんぴつを______。

7. (My) friend will give Ashton Kutcher a present.
友だちはアシュトン・カッチャーにプレゼントを______。

8. Lindsey Lohan will give (my) friend (some) food.
リンジー・ローハンは友だちに食べ物を______。

9. (My) friend will give Eva Longoria a letter.
友だちはエバ・ロンゴリアに手紙を______。

10. (My) sister will give Tanaka-san a bike.
妹 は田中さんに自転車を______。

11. Yamada-san will give (me) a CD.
山田さんは( 私 に) CDを______。

226
12. Suzuki-san will teach (my) mom Hiragana.
　鈴木さん母にひらがなを______。

13. Hayashi-san will give (my) sister tea.
　林さんは姉にお茶を______。

14. Ashton Kutcher will teach (my) friend how to do Twitter.
　アシュトン・カッチャーは友だちにTwitterの仕方を______。

15. Patrick Dempsey will give (my) friend a book.
　パトリック・デンプシーは友だちに本を______。

16. (I) will ask Ogawa-san (his) phone number.
　(私は) 小川さんに電話番号を______。

17. Whoopi Goldberg will give (my) friend a postcard.
　ウーピー・ゴールドバーグは友だちにはがきを______。

18. Ito-san will give (me) a movie.
　伊藤さんは(私に)映画を______。

19. Yoshida-san will teach (me) Japanese words.
　吉田さんは(私に)日本語の言葉を______。

20. (My) brother will give Sato-san a dog.
　兄は佐藤さんにいぬを______。

21. (I) will lend Hayashi-san a camera.
　(私は)林さんにカメラを______。

22. Kimura-san will ask (me) the time.
　木村さんは(私に)時間を______。

23. (My) friend will lend Whoopi Goldberg a bag.
　友だちはウーピー・ゴールドバーグにかばんを______。

24. (I) will give Yoshida-san a TV.
　(私は)吉田さんにテレビを______。
25. (My) father will teach Ito-san American politics.
父は伊藤さんにアメリカの政治を______。

26. (My) friend will teach Patrick Dempsey Kanji.
友だちはパトリック・デンプシーに漢字を______。

27. (My) friend will give Tom Cruise flowers.
友だちはトム・クルーズに花を______。

28. Sasaki-san will lend (me) money.
佐々木さんは( 私 に) お金を______。

29. Sato-san will lend (my) sister an umbrella.
佐藤さんは姉にかさを______。

30. (My) brother will lend Kato-san a dictionary.
兄は加藤さんに辞書を______。

31. (I) will give Kimura-san fruit.
( 私 は) 木村さんにくだものを______。

32. Ogawa-san will give (my) brother a cat.
小川さんは弟にねこを______。

33. Tanaka-san will ask (my) brother the meaning of a word.
田中さんは弟に言葉の意味を______。

34. Shimizu-san will give (my) dad a watch.
清水さんは父に時計を______。

35. Eva Longoria will ask (my) friend (his) name.
エバ・ロンゴリアは友だちに名前を______。

36. (My) sister will ask Yamada-san (her) address.
妹は山田さんに住所を______。
Answer sheet for written production posttest

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1. ___________________________</th>
<th>19. ___________________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td></td>
<td>2. ___________________________</td>
<td>20. ___________________________</td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td>3. ___________________________</td>
<td>21. ___________________________</td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td>4. ___________________________</td>
<td>22. ___________________________</td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td>5. ___________________________</td>
<td>23. ___________________________</td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td>6. ___________________________</td>
<td>24. ___________________________</td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td>7. ___________________________</td>
<td>25. ___________________________</td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td>8. ___________________________</td>
<td>26. ___________________________</td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td>9. ___________________________</td>
<td>27. ___________________________</td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td>10. ___________________________</td>
<td>28. ___________________________</td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td>11. ___________________________</td>
<td>29. ___________________________</td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td>12. ___________________________</td>
<td>30. ___________________________</td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td>13. ___________________________</td>
<td>31. ___________________________</td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td>14. ___________________________</td>
<td>32. ___________________________</td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td>15. ___________________________</td>
<td>33. ___________________________</td>
</tr>
<tr>
<td>16.</td>
<td></td>
<td>16. ___________________________</td>
<td>34. ___________________________</td>
</tr>
<tr>
<td>17.</td>
<td></td>
<td>17. ___________________________</td>
<td>35. ___________________________</td>
</tr>
<tr>
<td>18.</td>
<td></td>
<td>18. ___________________________</td>
<td>36. ___________________________</td>
</tr>
</tbody>
</table>
Appendix P
Written recognition posttest items

1. (I) will give Yamada-san a movie
   (私は) 山田さんに映画を________。

2. Patrick Dempsey will teach (my) friend a song.
   パトリック・デンプシーは友だちに歌を________。

3. (My) sister will give Hayashi-san a pen.
   姉は 林 さんにペンを________。

4. Sasaki-san will give (my) brother a dog.
   佐々木さんは兄にいぬを________。

5. Shimizu-san will teach (me) Japanese.
   清水さんは( 私 に)日本語を________。

6. Sato-san will give (me) a present.
   佐藤さんは( 私 に)プレゼントを________。

7. (My) friend will give Whoopi Goldberg tea.
   友だちはウピー・ゴールドバーグにお茶を________。

8. Suzuki-san will give (me) a TV.
   鈴木さんは( 私 に)テレビを________。

9. Whoopi Goldberg will lend (my) friend an umbrella.
   ウーピー・ゴールドバーグは友だちにかさを________。

10. (My) friend will give Patrick Dempsey a cat.
    友だちはパトリック・デンプシーにねこを________。

11. Ito-san will teach (my) dad Japanese politics.
    伊藤さんは父に日本の政治を________。

12. Tanaka-san will give (me) a radio.
    田中さんは( 私 に)ラジオを________。
13. Yamada-san will ask (my) sister the time.

山田さんは妹に時間を________。

14. (My) brother will ask Tanaka-san (his) birthday.

弟は田中さんに誕生日を________。

15. Yoshida-san will give (my) mom water.

吉田さんは母に水を________。

16. (My) friend will lend Tom Cruise a book.

友だちはトム・クルーズに本を________。

17. Kato-san will lend (my) brother a dictionary.

加藤さんは兄に辞書を________。

18. Tom Cruise will give (my) friend fruit.

トム・クルーズは友だちにくだものを________。

19. Kimura-san will give (my) sister a letter.

木村さんは妹に手紙を________。

20. Lindsey Lohan will ask (my) friend (his) phone number.

リンジー・ローハンは友だちに電話番号を________

21. (I) will ask Kimura-san (her) name.

(私は)木村さんに名前を________。

22. Hayashi-san will lend (me) a pencil.

林さんは(私に)えんぴつを________。

23. Eva Longoria will give (my) friend a bike.

エバ・ロンゴリアは友だちに自転車を________。

24. (My) mother will teach Suzuki-san English.

母は鈴木さんに英語を________。
25. (My) dad will give Shimizu-san bread.
父は清水さんにパンを_______。

26. (I) will teach Yoshida-san an English word.
( 私は) 吉田さんに英語の言葉を_______。

27. (I) will give Ito-san a book.
( 私は) 伊藤さんに本を_______。

28. (I) will lend Sasaki-san (my) bag.
( 私は) 佐々木さんにかばんを_______。

29. (My) sister will lend Sato-san a camera.
姉は佐藤さんにカメラを_______。

30. (My) friend will give Lindsey Lohan a postcard.
友だちはリンジー・ローハンにはがきを_______。

31. (My) brother will give Ogawa-san a CD.
弟は小川さんにCDを_______。

32. (My) friend will teach Ashton Kutcher Hiragana.
友だちはアシュトン・カッチャーにひらがなを_______。

33. (I) will give Kato-san a watch.
( 私は) 加藤さんに時計を_______。

34. (My) friend will ask Eva Longoria (her) address.
友だちはエバ・ロンゴリアに住所を_______。

35. Ashton Kutcher will give (my) friend flowers.
アシュトン・カッチャーは友だちに花を_______。

36. Ogawa-san will ask (me) the meaning of a word.
小川さんは( 私に) 言葉の意味を_______。
Answer sheet for written recognition posttest

<table>
<thead>
<tr>
<th>Word bank</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. すすめる</td>
<td>2. のこす</td>
<td>3. おしえる</td>
<td>4. いただく</td>
<td>5. あげる</td>
<td></td>
</tr>
<tr>
<td>6. ゆだねる</td>
<td>7. くれる</td>
<td>8. しょうかいする</td>
<td>9. かす</td>
<td>10. きく</td>
<td></td>
</tr>
</tbody>
</table>

| 1. ______________ | 19. ______________ |
| 2. ______________ | 20. ______________ |
| 3. ______________ | 21. ______________ |
| 4. ______________ | 22. ______________ |
| 5. ______________ | 23. ______________ |
| 6. ______________ | 24. ______________ |
| 7. ______________ | 25. ______________ |
| 8. ______________ | 26. ______________ |
| 9. ______________ | 27. ______________ |
| 10. ______________ | 28. ______________ |
| 11. ______________ | 29. ______________ |
| 12. ______________ | 30. ______________ |
| 13. ______________ | 31. ______________ |
| 14. ______________ | 32. ______________ |
| 15. ______________ | 33. ______________ |
| 16. ______________ | 34. ______________ |
| 17. ______________ | 35. ______________ |
| 18. ______________ | 36. ______________ |
Appendix Q
Questionnaire questions

Used for participants in the Planning Groups:

1. Did you find the activities which you did with the Japanese speaker to be easy or difficult?

   1  2  3  4  5
   very easy  very difficult

Please explain the reasons for your choice.

2. Was the planning you did before the activity helpful?

   1  2  3  4  5
   very helpful  not helpful at all

Please explain the reasons for your choice.

3. Do you have any other comments about the activities?

4. What do you think was the focus of this research project?

5. Is there anything you noticed while you were doing the activities?

6. Is there anything you noticed about what the Japanese speaker was saying to you?

7. Is there anything you noticed about the Japanese language?

(On the separate page)

8. If you had to explain the verbs ageru and kureru to a friend who didn’t know about them, what might you say? Please include any ideas you have, even if they’re vague and you’re not sure.
1. Did you find the activities which you did with the Japanese speaker to be easy or difficult?

1 2 3 4 5
very easy very difficult

Please explain the reasons for your choice.

2. Do you have any comments about the activities?

3. What do you think was the focus of this research project?

4. Is there anything you noticed while you were doing the activities?

5. Is there anything you noticed about what the Japanese speaker was saying to you?

6. Is there anything you noticed about the Japanese language?

(On the separate page)

7. If you had to explain the verbs ageru and kureru to a friend who didn’t know about them, what might you say? Please include any ideas you have, even if they’re vague and you’re not sure.
Appendix R

Noticing from four data sources

<table>
<thead>
<tr>
<th>Participants</th>
<th>Think-alouds</th>
<th>Stimulated recalls</th>
<th>Questionnaire</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td></td>
<td>Self Family In/Out-group</td>
<td>Self In/Out-Group</td>
<td></td>
</tr>
<tr>
<td>P17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P22</td>
<td></td>
<td>Self</td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>P23</td>
<td>Self</td>
<td>Self</td>
<td>Self</td>
<td>Self Family</td>
</tr>
<tr>
<td>P24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P33</td>
<td></td>
<td></td>
<td></td>
<td>Self</td>
</tr>
<tr>
<td>No Planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td>N/A</td>
<td></td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>P19</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P25</td>
<td>N/A</td>
<td></td>
<td>Family</td>
<td>Self</td>
</tr>
<tr>
<td>P26</td>
<td>N/A</td>
<td></td>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>P27</td>
<td>N/A</td>
<td></td>
<td>In/Out-group</td>
<td>In/Out-group</td>
</tr>
<tr>
<td>P29</td>
<td>N/A</td>
<td></td>
<td>Family</td>
<td>Family</td>
</tr>
<tr>
<td>P30</td>
<td>N/A</td>
<td></td>
<td></td>
<td>Self</td>
</tr>
<tr>
<td>P31</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P32</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P34</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


