Despite the frequency with which verbal reports are used in SLA to gather data on learners' cognitive processes (e.g., Bowles, 2003, 2004; Mackey, Gass, & McDonough, 2000; Rosa & Leow, 2004a, 2004b), only two studies (Bowles & Leow, 2005; Leow & Morgan-Short, 2004) have investigated verbal reports' reactivity (i.e., whether they alter cognitive processes) during second language (L2) reading. This is the first study to investigate the reactivity of verbal reports on a L2 problem-solving task. First-semester learners of Spanish were assigned to one of six experimental conditions, which differed in terms of the type of verbalization (i.e., metalinguistic, nonmetalinguistic, or silent) and the type of feedback (i.e., implicit vs. explicit). Results show that metalinguistic verbalization significantly increased time on task and also hindered participants' ability to produce exemplars of the target structure seen during the experimental task. However, neither type of verbalization significantly affected participants' ability to produce novel exemplars of the target structure, and there was no interaction between verbalization and feedback.
Verbal reports have been used increasingly in the past three decades to study cognitive processes, so much so that “both concurrent and retrospective verbal reports are now generally recognized as major sources of data on subjects’ cognitive processes in specific tasks” (Ericsson & Simon, 1993, p. xi). They are widely used in both first language (L1) and second language (L2) research to gain insight into subjects’ cognitive processing, thought processes, and strategies. Examples of verbal reports used in SLA research to gather data on learners’ cognitive processes during their interaction with the L2 data include concurrent think-alouds (e.g., Camps, 2003; Egi, 2004; Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999) and retrospective verbal reports, also known as stimulated recalls (e.g., Félix-Brasdefer, 2004; Mackey, Gass, & McDonough, 2000; Nabei & Swain, 2002).

VERBAL REPORTS

Use of Verbal Reports in L1 and L2 Research

Verbal reports have a long history in language research. Since the 1980s, verbal reports have been used systematically in L1 reading (Alvermann, 1984; Cohen, 1986, 1987; Earthman, 1992; Fehrenbach, 1991; Folger, 2001; Gordon, 1990; Harmon, 2000; McGuire & Yewchuk, 1996; Rosshenise & Meister, 1992), writing (Breetvelt, 1994; Durst, 1987; Green & Sutton, 2003; Zellermayer & Cohen, 1996), and language testing research (Anderson, Bachman, Perkins, & Cohen, 1991; Gavin, 1989; Wijgh, 1996).1 Similarly, introspective methods, including verbal reports, have been used extensively as a data elicitation technique in SLA. Since SLA was first studied systematically in the early 1970s, there has been some debate over whether learners’ introspections, or their production alone, should be used for theory building (Corder, 1973; Selinker, 1972). Although the debate over the use of verbal reports continues today, researchers in virtually all strands of SLA are using both concurrent and retrospective verbal reports increasingly, regardless of their theoretical positions, to investigate such phenomena as L2 reading and writing (Cavalcanti & Cohen, 1990; Cohen, 1987; Cohen & Cavalcanti, 1987; Hosenfeld, 1976, 1977, 1979, 1984), L1 and L2 strategies (Chamot & El Dinary, 1999; Davis & Bistodeau, 1993; Nevo, 1989; Yamashita, 2002), L2 test-taking strategies (Cohen, 2000; Norris, 1992; Warren, 1996), translation (Enkvist, 1995; Færch & Kasper, 1986; Jaaskelainen, 2000; Kern, 1994; Ronowicz, Hehir, Kaimi, Kojima, & Lee, 2005), interlanguage pragmatics (Cohen, 1998a, 1998b; Cohen & Hosenfeld, 1981; Cohen & Olsttain, 1993; Félix-Brasdefer, 2004; Kasper, 1999; Kasper & Blum-Kulka, 1993; Kasper & Rose, 2002), conversational interaction (Egi, 2004; Mackey et al., 2000; Nabei & Swain, 2002; Philp, 2003; Polio, Gass, & Chapin, 2006), noticing during interaction with L2 input (Bowles, 2003; Leow, 1997, 1998a, 1998b, 1999, 2000, 2001a, 2001b; Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999), and the relationship between explicit and implicit L2
knowledge (Ellis, 2004; Hu, 2002). In his recent article on the definition and measurement of explicit knowledge, Ellis stated that “collecting verbal explanations . . . would appear, on the face of it, to provide the most valid measure of a learner’s explicit knowledge” (p. 263). SLA researchers use verbal reports in an effort to gain information about learners’ cognitive processes as they interact with the L2. As Gass and Mackey (2000) stated, “understanding the source of second language production is problematic because often there are multiple explanations for production phenomena that can only be assessed by exploring the process phenomena” (p. 26). Verbal reports are one means of accessing such process phenomena.

**Categorization of Verbal Reports**

All verbal reports are not identical; rather, there are important differences among verbal reports based on the conditions under which they are collected. Ericsson and Simon (1984, 1993) proposed a typology to categorize verbal reports based on both the temporal frame in which they are collected and the level of detail of reporting. Concurrent reports are those collected as subjects verbalize while performing the task in question, whereas retrospective reports are collected when subjects verbalize some time after performing the task. In addition to categorizing verbal reports in terms of temporal space, Ericsson and Simon (1984, 1993) also distinguished between reports that require subjects to verbalize their thoughts per se and those that require subjects to verbalize additional information, such as explanations and justifications. Following from recent research by Bowles and Leow (2005), verbalizations of thoughts per se will be referred to as nonmetalinguistic and verbalizations of explanations and justifications will be referred to as metalinguistic.

**Validity of Verbal Reports**

Despite the frequency with which verbal reports have been used as a methodological tool to gain insight into L2 learners’ cognitive processes, systematic research on their validity in SLA is just beginning. Such research is clearly warranted, as Ericsson and Simon (1993), among others, have identified potential threats to the validity of both retrospective and concurrent verbal reports. For retrospective reports, because participants verbalize some time after completing a task, there is a potential for veridicality, or memory decay. In other words, retrospective verbal reports might not accurately reflect thought processes because participants simply might not recall what they were thinking as they completed the given task. However, this threat can be minimized if there is only a short delay between task performance and verbalization. Similarly, if learners are provided with some stimulus, such as a video or audiotape of their performance, as described in Gass and Mackey’s (2000) account of stimulated recalls, the possibility of veridicality is also lessened.
For concurrent verbal reports, the threat to validity is not veridicality because verbalization and task performance are concomitant. Rather, the validity of concurrent reports has been questioned because it is not known whether the act of verbalizing while completing a task is reactive, acting as an additional task and altering cognitive processes rather than providing a true reflection of thoughts, as Ellis (2001) and Jourdenais (2001) have suggested in the SLA literature. The current study investigates only concurrent verbal reports and therefore only the issue of reactivity, operationalized as an observed alteration of learners’ cognitive processes due to verbalization during task completion.

In addition to categorizing verbal reports, Ericsson and Simon’s (1993) model made predictions about their validity. Specifically, it predicted that nonmetalinguistic verbalizations (which Ericsson and Simon refer to as Type 1 verbalizations) would be largely nonreactive; that is, they would reflect the nature of cognitive processes fairly accurately, merely slowing processing slightly. Furthermore, the model predicted that metalinguistic verbalization (Type 2 and 3 verbalizations in Ericsson and Simon’s model) might be more reactive.

### INVESTIGATIONS INTO THE REACTIVITY OF VERBAL REPORTS

#### Previous Non-SLA Studies

Whereas investigations into the validity of verbal reports in SLA are just beginning, reactivity has been studied in the psychological literature on problem solving for more than 50 years. The earliest studies in this vein (Ewert & Lambert, 1932; Katona, 1940) investigated the effects of experimenter-provided verbalizations on participants’ task performance. In these studies, the experimenter would identify a series of principles necessary for successful completion of a problem-solving task and would then verbally state these essential principles to one group of participants, hypothesizing that the performance of the participants who received this verbalization would improve, whereas the performance of the participants who did not receive such verbal guidance would remain unchanged. By the 1950s, however, an ideological shift had begun to take place in psychology, whereby the focus of experimentation was redirected, moving away from the study of external modifications made to the input and toward participants’ own internally driven cognitive processes. It was at this point that researchers started to ask participants to verbalize their thoughts as they completed tasks and to investigate the reactivity of participants’ verbalizations on task performance (Brunk, Collister, Swift, & Stayton, 1958; Gagné & Smith, 1962; Hafner, 1957; Marks, 1951).

Over the years, as the use of verbal reports has increased, numerous studies—primarily in cognitive psychology—have investigated their potential reactivity. A synthesis of the studies that included verbalization as an independent variable indicates that nonmetalinguistic verbalization did not seem
to affect task performance (i.e., a finding of nonreactivity; Ericsson, 2002; Ericsson & Simon, 1993), which suggests that nonmetalinguistic verbalization might be a valid method of capturing internal thought processes.\(^2\) However, results indicated that, overall, verbalization tends to increase the overall solution time. Additional studies that investigated the reactivity of nonmetalinguistic reports but were not described in Ericsson and Simon (Biggs, Rosman, & Serfenian, 1993; Brinkman, 1993; Knoblich & Rhenius, 1995; Williams & Davids, 1997) found a similar lack of reactivity.

However, when participants are required to provide explanations or justifications (metalinguistic verbalizations), the opposite pattern of findings emerges. Most studies that examined the effects of metalinguistic verbal reports compared to a silent control group (Berry, 1983; Berry & Broadbent, 1984; Bower & King, 1967; Davis, Carey, Foxman, & Tarr, 1968; McGeorge & Burton, 1989; Short et al., 1991; Stanley, Mathews, Buss, & Kotler-Cope, 1989; Wilder & Harvey, 1971; Wilson & Schooler, 1991) found metalinguistic reports to be reactive; that is, provision of metalinguistic verbal reports seems to have an effect on task performance when compared to silent controls.\(^3\) In general, time on task (latency) was not reported, but Ericsson and Simon’s (1993) prediction that metalinguistic verbalization would require extra processing time in comparison with other types of verbalizations or with a control holds true for almost all of the studies that reported latency (Ahlum-Heath & di Vesta, 1986; Allwood, 1990; Carpenter, Just, & Schell, 1990; Fidler, 1983; Gagné & Smith, 1962; Robinson, 2001). These findings were also largely confirmed by a review of additional empirical studies examining the reactive effects of metalinguistic reports (Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995; Robinson; Schooler, Ohlsson, & Brooks, 1993).

Nevertheless, caution should be taken in interpreting these results, because most of the studies reviewed employed nonverbal tasks, which limits their relevance to SLA. In fact, of all the studies on reactivity, only six used verbal tasks (Lass, Klettke, Lüer, & Ruhlernder, 1991; Mathews et al., 1989; Rhenius & Deffner, 1990; Russo, Johnson, & Stephens, 1989; Short et al., 1991; Stratman & Hamp-Lyons, 1994). Furthermore, some research (e.g., Deffner, 1989) has provided evidence that “the causes of reactivity are not general but due jointly to the demands of the task and to verbalization” (Russo et al., p. 763).

Table 1 summarizes the five studies that expressly set out to investigate the effects of task type on reactivity. Three of the five (Russo et al., 1989; Short et al., 1991; Stratman & Hamp-Lyons, 1994) found differential effects based on the type of task, and, perhaps most relevant to the present study, Stratman and Hamp-Lyons found differential effects for thinking aloud on aspects of a verbal task (L1 writing). Although the sample size was too small to permit statistical analyses, the authors indicated that thinking aloud appeared to hinder participants’ ability to detect and fix organizational errors in the text but at the same time to improve their ability to detect pronoun errors. Based on these findings, it is clear that the effects of verbalization do not apply universally to all types of tasks.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Tasks</th>
<th>Reactivity</th>
<th>Nonreactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berardi-Coletta et al.</td>
<td>Undergraduates</td>
<td>Exp 1: Tower of Hanoi</td>
<td>Accuracy and time for both tasks</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exp 2: Katona card problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhenius &amp; Deffner</td>
<td>Undergraduates</td>
<td>1. Sentence assembly</td>
<td>Time for all tasks$^a$</td>
<td>Accuracy for all tasks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Anagrams</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Geometrical puzzles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Ravens matrices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russo et al. (1989)</td>
<td>Undergraduates</td>
<td>1. Mathematical task (mental addition)</td>
<td>Accuracy for gambles and addition</td>
<td>Accuracy for anagrams and ravens matrices; Time for all tasks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Verbal task (anagrams)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Numerical task (gambles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Pictorial task (Ravens matrices)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short et al. (1991)*</td>
<td>Fifth-grade students with differing Iqs</td>
<td>1. Spatial analogies</td>
<td>Accuracy for all analogies, but more reactive for spatial analogies with lower IQ students</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Verbal analogies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stratman &amp; Hamp-Lyons (1994)*</td>
<td>Undergraduate and graduate students</td>
<td>Revision of L1 text</td>
<td>Accuracy for both text features, but in different directions$^b$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Organizational features</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Word-level features</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Indicates that time on task was not reported.

$^b$Differences between solution times for think-aloud and silent groups were significant for all tasks except sentence assembly, for which they approached significance.

$^b$Thinking aloud appeared to hinder participants’ ability to detect and fix organizational errors in the text and, conversely, improve their ability to detect pronoun errors.
Previous SLA Studies

The only two SLA studies on reactivity investigated the effects of verbalization during L2 reading (Bowles & Leow, 2005; Leow & Morgan-Short, 2004). Leow and Morgan-Short conducted the first SLA study to empirically investigate the reactivity of verbal reports. In their study, 77 beginning-level Spanish learners read a text in Spanish, either silently or while verbalizing nonmetalinguistically. Posttests measured text comprehension as well as recognition and written production of formal imperative verbal morphology. Results showed nonreactivity for concurrent nonmetalinguistic verbalization: The two groups did not differ significantly on either text comprehension or posttest scores on the target structure. Because time on task was not one of the study’s dependent variables, latency was not addressed.

Expanding on Leow and Morgan-Short’s (2004) study, Bowles and Leow (2005) sought to investigate the reactivity of both metalinguistic and nonmetalinguistic reports on text comprehension and item and system learning of a syntactic structure, the pluperfect subjunctive. Participants were 45 fifth-semester Spanish learners who were randomly assigned to either a control group or to one of two verbalization groups (nonmetalinguistic or metalinguistic). All participants read a text that included tokens of the target structure and then completed a series of comprehension and written production tasks. The groups differed only in the verbalization instructions they received; participants in the control group were silent, whereas those in the nonmetalinguistic group were instructed to say whatever passed through their minds, and those in the metalinguistic group were instructed to verbalize what information they were looking at, what thoughts they were having about any piece of information, how they evaluated different pieces of information, and why. Similar to Leow and Morgan-Short, results indicated that, compared to a control group, nonmetalinguistic verbalization did not significantly affect either comprehension or written production of the target form. However, metalinguistic verbalization caused a significant decrement in text comprehension but no significant difference for production when compared to either the control or the nonmetalinguistic group. Furthermore, results indicated that both verbalization groups took significantly more time to read the text and complete the posttests than the control group, and no significant difference was found between the nonmetalinguistic and metalinguistic groups in terms of time on task.

Motivation for the Current Study

Although verbal reports have been used in a number of SLA studies and stand to provide invaluable insight into the language learning process, their validity as a methodological tool in SLA research is just starting to be examined. Two SLA studies (Bowles & Leow, 2005; Leow & Morgan-Short, 2004) have investi-
gated type of verbal report as an independent variable during L2 reading. Although these studies' findings certainly shed some light on the validity of verbal reports, future studies are warranted to investigate the reactivity of such reports on different types of tasks. Because verbal reports have already been used in SLA research during a range of L2 tasks, including picture drawing and spot-the-difference tasks in oral interaction research (Nabei & Swain, 2002; Philp, 2003), role-play (Félix-Brasdefer, 2004) and discourse completion tasks in interlanguage pragmatics (Robinson, 1991), and task-essential problem-solving tasks targeting morphological and syntactic L2 structures (Leow, 1997, 1998a, 1998b, 2000, 2001a, 2001b; Rosa & Leow, 2004a, 2004b; Rosa & O’Neill, 1999), investigations into reactivity might be even more crucially needed, given that several studies in cognitive psychology have found verbal reports to be reactive for some tasks but not for others (e.g., Rhenius & Deffner, 1990; Russo et al., 1989).

In addition to the potential effect of task type, another factor that has not been examined in studies on the reactivity of verbal reports in SLA is feedback. Some studies in cognitive psychology (Brehmer & Lindberg, 1971; Hagafors & Brehmer, 1983; Hammond & Summers, 1972) have observed the separate and combined effects of immediate explicit feedback and concurrent verbalization on task performance. Hagafors and Brehmer's study is of particular interest because it investigated the effects of metalinguistic verbalization and explicit feedback on performance of a cue-probability learning task. Sixty-four high school students were randomly assigned to one of four groups: (a) an explicit feedback/justification group, (b) an explicit feedback/no justification group, (c) a no feedback/justification group, or (d) a no feedback/no justification group. All participants completed a series of learning trials followed by a similar experimental trial. Results indicated that verbal justification (metalinguistic verbalization) had a significant facilitative effect on task performance when no feedback was provided. However, when explicit feedback was provided, no difference was found between justification and nonjustification groups (metalinguistic verbalization and silent control groups). In this case, metalinguistic verbalization was found to be reactive in the absence of explicit feedback but nonreactive in the presence of such feedback.

To date, no SLA studies have been conducted to determine whether there is an interaction between the provision of feedback and verbalization. Given the prevalence of tasks involving some type of feedback (whether oral or written) in SLA studies and the increasing use of verbal reports as a methodological tool to gain insight into learners’ cognitive processes and access to explicit L2 knowledge (Ellis, 2004, 2005; Gass & Mackey, 2000; Schmidt, 2001), research appears to be essential not only to ensure the validity of the measure but also to help SLA researchers make progress in defining and determining the role of explicit knowledge in SLA.5

Additionally, there is no conclusive evidence about whether verbalization affects time on task. Many studies in cognitive psychology neglect to report
times for their verbalization groups and, in SLA research, only Bowles and Leow (2005) have investigated that question for L2 reading only.

**The Present Study**

To address these limitations, this study investigates the methodological issue of reactivity in SLA by examining the effects of type of verbalization (nonmetalinguistic vs. metalinguistic vs. silent control) on a L2 problem-solving task. Additionally, it considers the possible interaction between verbalization and feedback that has been documented in cognitive psychology. Production of both previously encountered and novel exemplars of the target structure is examined to determine the effects of verbalization on both item and system learning (Bowles & Leow, 2005).

Specifically, the following research questions guided the study:

1. Does the type of verbalization (nonmetalinguistic vs. metalinguistic vs. silent control) produced during a L2 problem-solving task have an effect on beginning L2 learners’ immediate written production of previously encountered or novel exemplars of the target syntactic structure when either implicit or explicit feedback is provided?
2. Does the type of verbalization (nonmetalinguistic vs. metalinguistic vs. silent control) significantly affect the amount of time it takes to complete the task?

On the basis of these research questions, two hypotheses are posited. The first hypothesis states that the type of verbalization will not affect beginning L2 learners’ immediate written production of previously encountered or novel exemplars when either explicit or implicit feedback is provided (Bowles & Leow, 2005; Leow & Morgan-Short, 2004). The second hypothesis claims that the type of verbalization will significantly affect the amount of time it takes to complete the task, with both experimental groups taking significantly longer to complete the task than the silent control group (based on the overall findings of cognitive psychology studies on reactivity and on Bowles & Leow).

**METHOD**

**Participants**

The original pool of participants consisted of 325 students enrolled in first-semester Spanish courses at two private universities in the United States. Only monolingual English speakers who demonstrated that they were unfamiliar with the target structure at the outset of the experiment were included in the final sample.

Native language and previous language study were controlled to prevent interference from knowledge of other languages that have similar word order
patterns and dative experiencers as Spanish, and only participants who scored zero on the written production pretest were included in the final sample to ensure that all participants had a minimal amount of knowledge of the target structure prior to the study. Participants were also excluded for not completing all parts of the study, for not providing audible verbal reports (if they were assigned to a verbalization condition), or for having received outside exposure to the target structure between the pretest and delayed posttests (as reported on a debriefing questionnaire). Of the original pool of 325 participants, 131 were eliminated based on these criteria, yielding a final sample of 194 participants. These participants were randomly assigned to one of the six experimental conditions; these groups differed by verbalization condition (metalinguistic, nonmetalinguistic, or none) and feedback type (implicit vs. explicit), as shown in Table 2.

### Target Structure

The target linguistic structure is the Spanish syntactic construction of the psych verb *gustar* “to like,” which requires a dative experiencer obligatorily doubled by a dative clitic. In English, however, experiencers can only have the nominative or accusative case (Montrul, 1997, 1998, 2004). Compare (1) and (2):

(1) a. *A Juan le gustan los caramelos.*
   "Juan likes candies."

b. *Los caramelos le gustan a Juan.*
   "Juan likes candies."

(2) *Juan likes candies.*
   "Juan likes candies."

Psych verbs like *gustar* “to like” are an ideal structure to compare the effects of feedback and verbalization because SLA research (Cazzoli-Goeta, Masullo,
& Young-Scholten, 2003; LoCoco, 1987; Montrul, 1998; Van Patten & Cadierno, 1993; VanPatten & Sanz, 1995) has shown them to be particularly difficult for L1 speakers of English. Sanz (1999) suggested that these sorts of difficult structures are the most appropriate targets for feedback because learners are unlikely to acquire them easily and completely without such instructional interventions.

**Experimental Tasks**

The experimental tasks consisted of 24 three-dimensional computerized mazes for which the target structure was task-essential. Participants entered each maze and moved through it virtually from a first-person vantage point using the mouse. Spanish pronouns and verb forms were placed along the paths of each maze, and only by choosing the correct forms to complete a grammatical L2 sentence could learners exit the maze. Specifically, at each forking path in the maze, there were either two pronouns or two verbs from which learners could choose. One form was correct, and one was a distracter that beginning-level learners were likely to choose (based on Gascón’s, 1998, analysis of common learner errors with gustar-type verbs). When learners chose an ungrammatical form, they ran into a wall and had to choose another form. However, when they chose a grammatical form, the path continued and eventually led them to an exit. See Appendix A for a visual representation of the decision paths in the mazes.

**Feedback and Verbalization Condition**

In the implicit feedback condition, this intrinsic feedback is all the learners received. In the explicit feedback condition, learners received explicit grammatical information about the target structure (in English) whenever they chose any pronoun or verb in the maze, whether correct or incorrect. Specifically, this explicit information provided learners with (a) a statement of whether the choice was correct (Right!) or incorrect (Sorry!) and (b) an explanation of why the choice was correct or incorrect (i.e., a metalinguistic grammatical explanation). This maze format was chosen for several reasons. Because the study was designed to investigate the validity of verbal reports used with L2 problem-solving tasks, it was important that the task be similar to problem-solving tasks used in previous SLA studies that required participants to think aloud (cf. the multiple-choice jigsaw puzzle in Rosa & Leow, 2004a, 2004b). Furthermore, to address the relationship between feedback and reactivity, it was essential that both implicit and explicit feedback versions of the task be created. The maze provided intrinsic, implicit feedback seamlessly and enabled a comparison with explicit feedback.

Participants were randomly assigned to one of the three verbalization conditions: metalinguistic, nonmetalinguistic, or silent (control). All participants
received general task instructions, and those in the metalinguistic and non-metalinguistic groups were also given instructions for verbalization (Appendixes B and C).

**Pretests and Posttests**

A controlled written production test was designed to measure participants’ ability to produce the target structure. The test required participants to write Spanish equivalents of English sentences, as illustrated by the distracter item in (3) and the target item in (4):

(3) Pedro writes letters. _________________ cartas.

(4) Luis and Marco like baseball. _________________ el béisbol.

Vocabulary was carefully controlled and limited to proper names and common nouns that had been presented in the learners’ first-semester textbook. As a further safeguard, the last noun of each sentence was provided at the end of the response line to ensure that the test measured learners’ syntactic knowledge rather than their vocabulary.

Participants completed three different versions of the written production test: a pretest, an immediate posttest, and a delayed posttest. The pretest was the shortest, with 24 sentences, of which 12 required the target structure and 12 were distracters. Because the posttests needed to accommodate enough tokens of the target structure to enable a comparison of item and system learning, they were slightly longer, with 35 sentences, of which 24 required the target structure (12 exemplars from the experimental task and 12 novel exemplars) and 11 were distracters.

To allow for a comparison with the existing body of research on reactivity (Bowles & Leow, 2005; Ericsson & Simon, 1993; Leow & Morgan-Short, 2004), only immediate posttest scores are analyzed and reported here.

**Testing Procedure**

This study was completed in three phases. On the first day, participants completed the language background questionnaire and the written production pretest. Two days later, participants completed the experimental tasks and the immediate posttest. Finally, 3 weeks after the treatment, participants completed the delayed posttest and a debriefing questionnaire that informed them of the target of the study and asked them to report any outside exposure to the target form during the period of the experiment.

On the day of the treatment, participants reported to the language laboratory in intact classes but were randomly assigned to one of the six experimental conditions (see Table 2) when they logged on to the testing computers. All
participants completed a familiarization session before starting the computerized experimental tasks. Additionally, participants in the two verbalization conditions received instructions on how to verbalize and then watched a demonstration in which a verbal report (either metalinguistic or nonmetalinguistic, according to condition) played as the cursor moved through a sample maze. Finally, the participants in the verbal reports groups were given the opportunity to warm up by verbalizing while working through one final sample maze. To ensure that participants in all conditions received the same amount of input on the target structure over the course of the treatment, none of the sample mazes in this session contained the target form.

Immediately after the familiarization session, participants completed the experimental treatment on their computers by working through the 24 mazes at their own pace, with either implicit or explicit feedback, according to their verbalization condition. Participants’ mouse clicks (detailing the paths that they chose) and time on task were collected and stored in a database for subsequent scoring, coding, and analysis, as were their immediate posttest responses.

**Scoring Procedure**

Because the target structure was a complex syntactic construction, all tests were scored to allow for partial knowledge. For each item, learners received 1 point for (a) using the preposition *a* to prepose the experiencer, (b) using the correct clitic pronoun, and (c) correctly conjugating the verb *gustar* “to like.” Therefore, the maximum possible score on the pretest was 36 points, and the maximum possible score on the posttest was 72 (36 points for previously encountered exemplars and 36 for novel exemplars).

Because one independent variable was the type of verbalization, it was essential that all of the participants’ verbalizations be coded accurately as metalinguistic or nonmetalinguistic. To ensure that participants had verbalized according to the initial instructions they received, the researcher listened to all the participants’ verbal reports and coded them in a binary fashion (Bowles & Leow, 2005; Rosa & O’Neill, 1999). The most important feature that distinguishes the metalinguistic group from the nonmetalinguistic group is the provision of justifications. Participants in the metalinguistic condition should have provided a justification or explanation for each path they chose to follow in the maze. Because participants were not likely to go down every path in every maze, it was clear that all participants would not have the same number of justifications to make. To resolve this, the researcher kept track of (a) the paths each participant chose throughout the course of the 24 experimental mazes and (b) whether the participant provided a justification for each path. Despite the fact that manipulation checks were conducted as participants completed the tasks, each participant in the metalinguistic condition might not have provided a justification for every path
chosen. If a participant provided justifications for 50% or more of the paths chosen, his or her report was coded as metalinguistic. Conversely, if a participant provided justifications for fewer than 50% of the paths chosen, his or her report was coded as nonmetalinguistic. In practice, however, participants were never near this 50% cutoff. The verbal reports coded as nonmetalinguistic contained justifications for 0–20% of the paths chosen, whereas those coded as metalinguistic contained justifications for 70–100% of the paths.

Of the 125 total verbal reports, only 8 were reclassified, which indicates that overall participants verbalized according to the instructions they received. As a check on reliability, 20% of the verbal reports \((n = 25)\) were selected at random to be coded by a second coder. Intercoder reliability, calculated using Cohen’s kappa, was found to be high \((\kappa = .90)\).

Excerpts from both types of verbal reports are provided in (5) and (6). Note that these excerpts are from participants’ verbal reports as they worked on the 12th maze, in which the target sentence was “Ana le gustan las películas” “Ana likes movies.” In these excerpts, participants are at the final forking path in the maze, where they have the choice between the third person singular verb form *gusta* and the third person plural form *gustan*. (See Appendix A for a detailed description of this maze.)

(5) Excerpt from a nonmetalinguistic verbal report
It should be *gustan* and I really hope there’s not gonna be any wall here. [Selects the correct form, *gustan*, and receives explicit feedback.] Yes, I’m right! Takes me straight to the exit. Right! [Participant exits the maze.]

(6) Excerpt from a metalinguistic verbal report
So the *ella* [third person singular] form of *gustar* must be *gust- gusta* because you always add an –a to the end so choosing *gusta*. [Selects the incorrect form, *gusta*, and receives explicit feedback.] No? I guess I got it wrong. OK, it [the feedback] says that the verb agrees with the thing that is liked or *películas*. So it’s gonna be *gust-an* ’cause there’s more than one movie. [Selects *gustan*.] That’s it! [Participant exits the maze.]

**RESULTS**

Prior to conducting statistical analyses to address the research questions, reliability analyses were performed on both the pretest and immediate posttest. Reliability coefficients, computed using Cronbach’s alpha, were high on both tests: .92 for the pretest and .97 for the immediate posttest.⁸

**Research Question 1**

To determine whether the verbalization condition had a significant effect on participants’ immediate production of previously encountered or novel exemplars of the target structure when either implicit or explicit feedback was provided, a \(3 \times 2\) factorial ANOVA with verbalization condition (metalinguistic,
nonmetalinguistic, or silent) and type of feedback (implicit or explicit) as between-subjects factors was performed.

**Previously Encountered Exemplars.** Complete descriptive statistics on previously encountered exemplars are provided in Table 3. Results of the $3 \times 2$ factorial ANOVA revealed a significant main effect for verbalization condition, $F(2, 193) = 3.778, p < .05$, and a significant main effect for feedback, $F(1, 193) = 24.925, p < .0001$. No significant interaction was found between verbalization condition and feedback, $F(2, 193) = 1.284, p = .279$.

A post hoc Scheffé test was performed to identify which contrasts contributed to the main effect for verbalization. Results revealed a difference between the control and metalinguistic groups ($p < .05$), with the control group performing significantly better on the production of previously encountered exemplars than the metalinguistic group. However, no significant difference was found between the two experimental groups on previously encountered exemplars ($p = .162$) or between the control and nonmetalinguistic groups ($p = .869$). Additionally, an independent samples $t$ test was performed to further examine the main effect for feedback. Results indicated that explicit feedback was significantly more effective than implicit feedback, $t(192) = 6.229, p < .0001$.

**Novel Exemplars.** To determine whether verbalization condition had a significant effect on participants’ immediate production of novel exemplars of the target structure when either implicit or explicit feedback was provided, a separate $3 \times 2$ factorial ANOVA was performed. Complete descriptive statistics on novel exemplars are provided in Table 3.

Results of the $3 \times 2$ factorial ANOVA on novel exemplars revealed a significant main effect for feedback, $F(1, 193) = 29.438, p < .0001$, but no significant main effect for verbalization, $F(2, 193) = 1.713, p = .183$, or significant inter-
action between verbalization condition and feedback, $F(1, 193) = 2.384, p = .095$. The results, then, indicate that the verbalization condition did not affect learners’ ability to produce novel exemplars of the target structure. Again, as was the case with previously encountered exemplars, results of an independent samples $t$ test indicated that explicit feedback was significantly more effective than implicit feedback, $t(192) = 6.993, p < .0001$.

To summarize, main effects for verbalization condition and for feedback were found for the production of previously encountered exemplars. Regardless of verbalization condition, participants who received explicit feedback outperformed those who received implicit feedback, and, regardless of feedback condition, participants who verbalized metalinguistically performed significantly worse on the production of previously encountered exemplars than those who performed the tasks silently. However, a significant main effect for feedback was found only for the production of novel exemplars. There was no significant main effect for verbalization condition and no interaction between the two independent variables. In other words, only feedback had a significant effect on participants’ ability to extrapolate the rules for the target structure to items in new contexts. Verbalization condition had no effect (i.e., it was not reactive).

It was hypothesized that the type of verbalization would not affect beginning L2 learners’ immediate written production of previously encountered or novel exemplars when either explicit or implicit feedback was provided. The results partially confirm this hypothesis. Verbalization did not have an effect on participants’ ability to produce novel exemplars of the target structure. However, it affected participants’ ability to produce exemplars of the target structure that they had seen in the experimental tasks. Specifically, participants who verbalized justifications were able to produce those exemplars significantly less accurately than those who completed the tasks silently.

**Research Question 2**

To determine whether the type of verbalization had a significant effect on the amount of time it took participants to complete the experimental task, time spent on task was submitted to a one-way ANOVA. Mean times were 2043.67 s (34.06 min) for the control group, 2081.45 s (34.69 min) for the nonmetalinguistic group, and 2397.70 s (39.96 min) for the metalinguistic group. Results of the ANOVA showed a significant main effect for type of verbalization, $F(2, 193) = 15.763, p < .0001$. Post hoc Scheffé tests indicated that this main effect was due to significant differences between the metalinguistic group, on one hand, and the control and nonmetalinguistic groups, on the other hand. Effect sizes calculated for the control versus metalinguistic and nonmetalinguistic versus metalinguistic comparison were large ($d = 0.86$ and 0.88, respectively). In other words, the metalinguistic group spent significantly more time on the experimental task than the control or nonmetalinguistic groups. However, the con-
control and nonmetalinguistic groups spent a similar amount of time on task ($p = .860$).

It was hypothesized that the type of verbalization would significantly affect the amount of time it takes to complete the task, with both experimental groups taking significantly longer to complete the task than the silent control group. The results partially confirm this hypothesis. The type of verbalization significantly affected the amount of time it took to complete the task. However, the experimental groups did not spend significantly more time on task than the control groups. Only the metalinguistic group took significantly more time to complete the task; there was no significant difference between the control and nonmetalinguistic groups in terms of time on task.

**DISCUSSION**

The first research question investigated whether the verbalization condition had a significant effect on participants’ immediate production of previously encountered or novel exemplars of the target structure when either implicit or explicit feedback was provided. With regard to previously encountered exemplars, this question was answered affirmatively because verbalization of justifications appears to have had a detrimental effect on participants’ ability to produce exemplars of the target structure seen in the experimental tasks. Therefore, verbalizing metalinguistically while completing the tasks appeared to have interfered with participants’ item learning ability. However, the finding of nonreactivity for verbalization with regard to production of novel exemplars of the target structure tempers this finding, suggesting that participants, regardless of verbalization group, were able to extrapolate what they learned in the tasks to correctly produce the structure in new contexts on the posttest. This suggests, therefore, that verbalization did not interfere with participants’ system learning. Furthermore, in neither case was there an interaction between verbalization condition and feedback.

These findings are not surprising in light of the non-SLA literature on the reactivity of verbal reports. Specifically, Ericsson and Simon’s (1993) model predicted that verbalization of thoughts per se, without the requirement to verbalize justifications, should provide a fairly pure reflection of thought processes and should therefore not be reactive. Their model specified that this is especially true when the verbalizations produced in nonmetalinguistic conditions would have been generated as a normal part of the thought process.

With regard to metalinguistic verbalization, Ericsson and Simon’s (1993) model predicted that verbalizations that require participants to verbalize additional specific information, such as explanations or justifications that might not otherwise be generated as a part of the thought process, will be reactive and will affect task performance. It is not surprising, then, given these predictions and the findings of non-SLA studies, that metalinguistic verbalization had a significant effect on participants’ immediate production of previously encoun-
tered exemplars. It seems that having to provide justifications for their decision path choices (Appendix A) interfered with participants’ ability to produce those same sentences on the posttest.

In this study, neither a significant main effect for verbalization nor a significant interaction between verbalization and feedback was found for the production of novel exemplars of the target structure: There was a lack of reactivity. Therefore, neither type of verbalization seems to have had an effect on system learning, as measured by the posttest.

Compared to the two SLA studies (Bowles & Leow, 2005; Leow & Morgan-Short, 2004) that have examined the reactivity of verbalization during a L2 reading task, this is the first study to investigate the reactivity of verbalization during a L2 problem-solving task. Despite this difference in tasks, a similar pattern of results was obtained, with no effect for either nonmetalinguistic or metalinguistic verbalization on the immediate production of novel exemplars of the target structure.

Moreover, no significant interaction between verbalization condition and type of feedback (explicit vs. implicit) was found, in contrast with studies from cognitive psychology (e.g., Hagafors & Brehmer, 1983) that examined this interaction. Hagafors and Brehmer found that when no feedback was provided, participants in a metalinguistic verbalization group performed significantly better on the experimental task than those in a nonmetalinguistic verbalization condition. However, when explicit feedback was provided, there was no significant difference in task performance for participants in the two verbalization conditions. One possible explanation for the different findings is the type of task involved. Hagafors and Brehmer used a cue-probability learning task in which participants had to predict a value based on a number of probabilistic cues given over a series of trials. In this type of task, participants could consistently give mistaken weights to the cues, trial after trial, and still be able to finish the task. However, here, participants interacted with tasks in which the L2 grammar structure was task-essential (Loschky & Bley-Vroman, 1993): They could only complete the task (progress from one maze to another) by forming grammatically correct sentences in Spanish. Perhaps for this reason, a significant interaction between feedback and verbalization was not found.

The results also indicate that the type of verbalization significantly affected the amount of time it took to complete the task. Specifically, participants in the metalinguistic group took significantly longer to complete the experimental task than did participants in either the silent control or the nonmetalinguistic groups. Ericsson and Simon’s (1993) model predicted that, overall, verbalization tends to be reactive for latency because the additional time needed for verbalization increases the overall solution time. This prediction is partially confirmed here, because only the metalinguistic group spent significantly more time on task than the other two groups. This finding is also consistent with those of non-SLA studies as well: Although many of the studies did not report latency, 10 of the 12 studies (83%) that reported latency found it to be reactive for metalinguistic reports, thus supporting Ericsson

This study’s results contradict those of Bowles and Leow (2005), who found that participants in both metalinguistic and nonmetalinguistic verbalization groups took significantly longer to complete the L2 reading task than did participants in the silent control group. Here, only the metalinguistic group was found to be reactive for latency. However, the findings can be reconciled because of the differences in the experimental tasks used: Whereas Bowles and Leow examined the reactivity of verbalization on a L2 reading task, this study investigated reactivity on a L2 problem-solving task. It is logical that a silent control group would be able to read a passage and answer related questions more quickly than either verbalization group. However, in a L2 problem-solving task for which the reading load is not heavy, it seems plausible that only the provision of justifications would significantly increase time on task. Given the differing findings with regard to latency of nonmetalinguistic reports, the question of how verbalization affects time on task remains somewhat of an open issue.

**LIMITATIONS AND FUTURE RESEARCH**

**Limitations**

Because this is the first study to investigate the reactivity of verbal reports on a L2 task other than reading, care should be taken overall in interpreting the findings. Additionally, a number of specific limitations are pertinent.

A first limitation concerns the pool of participants, who were first-semester L2 learners unfamiliar with the target structure at the outset of the experiment. It is not clear what effect learners’ proficiency might have had on reactivity and whether learners who had some degree of prior knowledge of the target structure might have performed differently given the requirement to verbalize during the task.

Similarly, the target syntactic structure was task-essential (Loschky & Bley-Vroman, 1993), so learners could not successfully complete the experimental tasks without using the target grammatical structure correctly. Certainly, the task-essentialness of the structure could have had an effect on the results, because the task pushed learners to attend to the grammar as they worked through the experimental tasks.

Finally, just one type of posttest (a highly controlled written production test) was used. This can be seen as a benefit in the sense that such a controlled test enables direct comparison among participants with little individual variation in possible correct responses. However, it can also be viewed as a limitation because no evidence about the potentially reactive effects of verbalization on more open-ended tasks was provided here. For
this reason, the results cannot speak to reactivity on less controlled tests of L2 knowledge.

**Future Research**

The present study contributes to a growing body of research investigating the validity of verbal reports as a SLA research method. Future research is certainly warranted in this area, as verbal reports continue to be used in various strands of L1 acquisition and SLA research. Especially important is an investigation into the reactivity of verbal reports with other types of SLA tasks.

Verbal reports could provide invaluable information to SLA researchers about language learners’ cognitive processes, information that would otherwise be unavailable to them. If these measures are shown to be valid and to accurately represent learners’ mental processes, they will greatly contribute to our understanding of L2 learning. Metalinguistic reports are of particular interest because, if they are found to be valid, they would allow researchers to gain even more insight into learners’ processing and justifications as they process and produce the L2. Data from these protocols would therefore allow researchers to target specifically the areas where students are in need of improvement and to address the faulty reasoning that is behind (at least some) L2 errors. This could prove to be one part of the solution to the perennial problem of identifying, understanding, and addressing L1 transfer errors such as those reported in Spada and Lightbown (1999). Some studies (Butler, 2002) have already shown that adult L2 learners seem to be able to consistently use their explicit knowledge to provide justifications of aspects of their L2 use (although these are, of course, not always correct). Further research on methods of measuring attention and awareness can only serve to move the field forward and enable us to learn more about our students’ processes as they interact with a L2 or foreign language.

It will be crucial to identify a systematic way to investigate reactivity and to determine which types of tasks, under which conditions, are likely to be reactive. Thus, researchers can begin to formulate principled guidance to those interested in using verbal reports with L2 learners. Only by synthesizing findings from previous SLA and cognitive psychology research and by focusing on issues and task features, such as task complexity (e.g., Deffner, 1989; Russo et al., 1989), already identified to play a role in reactivity, can researchers develop well-informed models. Once created, these models can be tested empirically using orthogonal pairings of L2 tasks with varying task demands and participants at varying levels of language proficiency to assess the separate and combined effects of language and task demands on reactivity.

Additionally, future research in the area of introspective SLA research methods is essential to furthering our understanding of the nature and use of explicit
knowledge in SLA (Ellis, 2004). Clearly, explicit knowledge cannot be measured directly, but verbal reports are one option that SLA researchers have for getting a window into their learners’ explicit knowledge, cloudy as the pane of glass may be. By collecting and carefully coding and analyzing verbal reports, we can gain insight into “what kinds of explicit knowledge learners exploit and in what ways” (Ellis, p. 268). Regardless of one’s theoretical position, as our objective tools for measuring explicit knowledge become sharper, verbal reports can be used as a complement, given that “another entirely different approach based on tasks that elicit verbal reports is [also] needed” (Ellis, p. 268).

(Received 7 June 2007)

NOTES

1. For a review of verbal reports used in language testing, see Green (1998).

2. Task performance has been used for more than 50 years in cognitive psychology (and in SLA since Leow & Morgan-Short’s, 2004, study) as an indicator of cognitive processing. In reactivity research, if silent groups and verbalization groups perform similarly on a task, it is assumed that verbalization did not significantly affect cognitive processes (and that the verbalizations produced are representative of the processing being done by the silent group). Nevertheless, as an anonymous SSLA reviewer suggested, task performance is far from being a foolproof indicator of cognitive processing. Indeed, it is possible that task performance is affected, but underlying cognitive processes are broadly unaffected, or, conversely, that task performance is not affected, when in fact verbalization has altered cognitive processes. Hopefully, advances in cognitive science and neuroimaging will enable us to probe the relationship between task performance and cognition more directly in the future.

3. This effect can be either positive (facilitating or improving task performance) or negative (hindering task performance).

4. In this study, production of both old and new exemplars was investigated to examine (a) the extent to which participants had simply memorized instances of the target structure that they had seen in the experimental passage or (b) to what extent they could formulate rules and extrapolate knowledge of the structure to new contexts. This is based on the distinction in psychology between two dissociable types of learning: item learning and system learning, or instance learning and rule learning (Shanks & St. John, 1994). Examination of both is important, because it has been shown that the stage of item learning (memorization of instances and fragments) occurs sequentially before the stage of system learning (creation and application of a rule) (Cruttenden, 1981).

5. An anonymous SSLA reviewer questioned whether there might be a relationship between explicit L2 knowledge and metalinguistic reporting, and, conversely, between implicit knowledge and non-metalinguistic reporting. The issue of defining and measuring explicit knowledge is certainly muddy, as recently pointed out by Ellis (2004, 2005). At present, I do not believe that there is enough empirical evidence to conclusively answer the question, but future SLA research in this area is certainly warranted.

However, I agree, following Ellis (2004), that explicit knowledge is conscious and, following Allport (1988), that what is verbalizable is conscious. By extension, it would seem that any type of verbal report would tap explicit knowledge to some extent. Nevertheless, the verbal reports that both Ellis and Hu (2002) described as possible measures of explicit L2 knowledge appear to be metalinguistic because they require learners to “explain grammatical features” (Ellis, p. 263).

6. Participants not only received verbalization instructions but also watched a short demonstration of a sample verbal report (either metalinguistic or nonmetalinguistic, according to their verbalization condition). This step is a methodological improvement over the two previous verbalization studies (Bowles & Leow, 2005; Leow & Morgan-Short, 2004) in that it follows Ericsson and Simon’s (1984, 1993) advice that participants should be provided with models of verbal reports in addition
to a short warm-up session during which they are allowed to practice verbalizing before the actual experiment begins.

7. The coding scheme described here is a methodological improvement over the one used in Bowles and Leow (2005), and the level of detail provided should allow greater replicability. Nevertheless, as an anonymous SSLA reviewer indicated, a more fine-grained coding scheme might reveal even more about the relationship between learners’ cognitive processes and reactivity.

8. A reliability coefficient $r$ of .7 or greater is considered high (Cook, 1998).

REFERENCES


Task Type and Reactivity


APPENDIX A

SAMPLE MAZE DECISION PATHS

Common learner errors addressed (Gascón, 1998)

ERROR TYPE 1

Omission of the preposition a
*Aña le gustan las películas.

DECISION TYPE 1

Decision types

Bare experiencer
*Aña

OR

σ +
experiencer
A Aña

ERROR TYPE 2

Omission of the obligatory dative clitic or incorrect clitic
*A Aña gustan las películas.
*A Aña les gustan las películas.

DECISION TYPE 2

Decision types

Incorrect clitic
*les

OR

Correct clitic
le

ERROR TYPE 3

Verb agreement with experiencer instead of with theme
*A Aña le gusta las películas.

DECISION TYPE 3

Decision types

Verb agreeing with experiencer
* gusta

OR

Verb agreeing with theme
gustan

Ins películas TO EXIT
APPENDIX B

VERBALIZATION INSTRUCTIONS (METALINGUISTIC CONDITION)

Instructions for Experimental Task

In this experiment I am interested in what you think about when you complete these tasks. In order to find out, I am going to ask you to THINK ALOUD as you work through the mazes. What I mean by “think aloud” is that I want you to verbalize your thoughts the entire time you are working on the tasks. I would like you to talk CONSTANTLY. Verbalize your thoughts as thoroughly as possible, describing what is happening on the screen and what thoughts you are having as you work through the maze. As you go through the maze, give a reason for choosing each path. Give a justification out loud for choosing each path and explain what you are thinking. Please provide as thorough a justification as possible, as if you were explaining to someone learning Spanish why you are choosing the paths you choose. Keep talking throughout and talk clearly into the microphone. You may speak in English.

Instructions for Written Posttest

Next you will complete a series of fill-in-the-blank exercises. Again, continue to think aloud as you complete these exercises. As you fill in the blanks, remember to give a justification out loud for your answers and explain what you are thinking. Please provide as thorough a justification as possible, as if you were explaining to someone learning Spanish why you filled in each blank the way you did. You may speak in English.

APPENDIX C

VERBALIZATION INSTRUCTIONS (NONMETALINGUISTIC CONDITION)

Instructions for Experimental Task

In this experiment I am interested in what you think about when you complete these tasks. In order to find out, I am going to ask you to THINK ALOUD as you work through the mazes. What I mean by “think aloud” is that I want you to verbalize your thoughts the entire time you are working on the tasks. I would like you to talk CONSTANTLY. Do not plan out what you are saying or explain what you’re saying. Just act as if you are alone in the room talking to yourself while you complete the tasks. What is most important is that you keep talking throughout and talk clearly into the microphone. You may speak in English. Just say whatever passes through your mind as you complete the tasks.
Instructions for Written Posttest

Next you will complete a series of fill-in-the-blank exercises. Again, continue to think aloud as you complete these exercises. Say whatever passes through your mind. You may speak in English.