



## **Effects of Explicit Instruction on Incidental Noticing of Metaphorical Word Sequences during a Subsequent Reading Task**

CHARLES MARK MUELLER\*

*University of Maryland and George Washington University*

Received: 14 June 2010 / Accepted: 20 July 2010

### **ABSTRACT**

Two experiments were conducted to determine whether explicit instruction focusing on metaphorical collocations would promote the incidental noticing of similar phrases by English learners during a subsequent reading task. Noticing was operationalized using the remember-know protocol and learning was measured on a fill-in-the-blanks test. In Experiment 1 ( $N = 36$ ), within-subjects comparisons showed that explicit instruction led to more incidents of noticing ( $p < .001$ ). Experiment 2 ( $N = 24$ ) sought to extend the findings by examining whether self-referential questions during instruction would lead to greater noticing compared to depersonalized questions. Experiment 2 confirmed the first experiment's finding of effects of explicit instruction on noticing, but failed to show significant effects for self-referential prompts on subsequent noticing, although such prompts did lead learners to produce higher word counts during the writing task. The two experiments suggest that explicit instruction promotes incidental noticing of important semantic contrasts in subsequent input.

### **KEY WORDS:**

Noticing, depth-of- processing, remember-know, self-referential, episodic memory, explicit, writing, reading, incidental, task-induced involvement

### **RESUMEN**

Este estudio presenta resultados de dos experimentos realizados con el objetivo de determinar si la instrucción explícita de colocaciones contribuye a que aprendices de inglés se percaten implícitamente de frases similares en una prueba de lectura posterior. El concepto de percatarse fue operacionalizado por medio del paradigma 'remember-know' mientras que el aprendizaje fue medido por medio de un ejercicio de 'fill-in-the-blanks'. En el primer experimento la comparación de variable intra-grupo reveló que los participantes efectivamente se percataron de un mayor número de frases ( $p < .001$ ). El segundo experimento, cuyo objetivo era obtener resultados análogos, examinó qué tipo de preguntas, reflexivas o impersonales, contribuiría a que los aprendices se percataran implícitamente de más colocaciones similares a las usadas en los estímulos. Los resultados se vieron corroborados, aunque los resultados obtenidos con preguntas reflexivas no alcanzaron significatividad estadística. Estos resultados sugieren que la instrucción explícita fomenta que los aprendices se percaten de contrastes semánticos importantes en el 'input'.

### **PALABRAS CLAVE:**

Toma de conciencia, intensidad del procesamiento, recordar-cómo, autoreferencial, memoria episódica, explícito, escribir, leer, secundario, implicación inducida por la tarea.

---

\**Address for correspondence:* Charles Mark Mueller. University of Maryland and George Washington University. 5700 Chapman Mill Dr. #210 Rockville, MD 20852 Ph. +1 443-631-7780. E-mail: [cmueller@umd.edu](mailto:cmueller@umd.edu)

## I. INTRODUCTION

A recurring debate within the field of second language acquisition (SLA) concerns the role of explicit and implicit learning. While disagreement remains regarding the relative efficacy of the two types of learning and their relationship with learner variables and particular linguistic features, the SLA field seems to be moving away from Krashen's (1981) strong view that explicit learning is of only minor importance to SLA (See DeKeyser, 2003; Norris & Ortega, 2000). Even so, the precise cognitive mechanisms and processes that underlie the effectiveness of explicit learning are still not fully understood.

### I.1. Review of the Literature

This paper presents two experiments that shed light on short-term effects of explicit instruction related to noticing. To clarify the background and motivation for the experiments, the paper begins with a brief review of SLA literature on the role of attention and its operationalization within experimental research.

#### *I.1.1. Explicit Instruction as a Trigger for Noticing*

Many researchers (DeKeyser, 2003; Doughty, 2001; Long, 1991; VanPatten, 1996) have expressed reservations regarding instructional practices that do not, at least occasionally, draw learners' attention to linguistic form. These concerns stem largely from the observation that learners often fail to notice forms in the input despite repeated exposure (Harley & Swain, 1984; Pica, 1983; Schmidt, 1983).

One proposed solution has been to use explicit instruction to target moderately difficult grammatical rules or semantic patterns. A possible benefit of explicit instruction may be its ability to trigger the incidental noticing of form-meaning connections within subsequent input (DeKeyser, 2003). In other words, learners may have a natural predilection to devote more processing resources to novel semantic and syntactic patterns that have recently been overtly targeted in instruction.

#### *I.1.2. Noticing in SLA*

Attention has been discussed extensively in SLA. Tomlin and Villa (1994) divide attention into three components (i.e., alertness, orientation, and detection), claiming that detection makes input available for further processing and thus provides a sufficient basis for learning. They (1994: 192) define *detection* as "the process that selects, or engages, a particular and specific bit of information." *Awareness*, on the other hand, is said to involve a specific

subjective experience that is marked by a cognitive or behavioral change, and by the ability to report or demonstrate an experience directly (Tomlin & Villa, 1994: 193).

Schmidt (1994) has argued that *noticing* (“awareness” in Tomlin and Villa’s model) is necessary to convert input to intake. For Schmidt (2001: 5), noticing merely requires attention to “elements of the surface structure of utterances in the input.” Following Schmidt (2001: 18), this paper defines *noticing* as “detection within focal attention accompanied by awareness.” It will be further stipulated (again, following Schmidt, 2001) that noticing typically involves a greater degree of control but does not necessarily entail metalinguistic understanding.

### ***1.1.3. Methodologies for Studying Noticing***

Due to its transitory nature, noticing is difficult to measure. Within the field of SLA, it has been studied through online measures (Bishop, 2004; Leow, 1998; Rosa & O’Neill, 1999), offline measures (Kim, 1995; Schmidt & Frota, 1986), and combined online and offline measures (Leow, 2001; Mackey, 2006).

#### ***1.1.3.a. Online Measures***

Online reports involve behavioral measures taken while a learner is engaged in a task. Such reports have some advantages over offline measures as they reflect mental processes that may be unavailable for recall at a later time. Leow in particular has extensively employed online reports in the form of think-aloud protocols. In a study that attempted to operationalize Tomlin and Villa’s components of attention, Leow (1998) gave four groups of Spanish learners four sets of crossword puzzles that were designed to facilitate (1) alertness, (2) alertness and orientation, (3) alertness, orientation, and detection, and (4) alertness and detection. Participants followed a verbal think-aloud protocol as they completed experimental tasks. He found significantly higher scores for the last two groups, who had detected the target forms. Leow’s study represents an interesting attempt to design tasks that elicit specific levels of awareness.

Online measures would seem to be ideal as they tap into cognitive processes as they occur; however, a potential problem with such measures is reactivity. Online measures can potentially introduce a confounding variable by fundamentally altering the way in which learners perform a task. Addressing this issue, Leow and Morgan-Short (2004) have presented evidence in favor of online verbal protocols in a study showing that reactivity did not affect the written performance of beginning Spanish learners who were performing a reading task. Yet it is still not clear whether online measures of tasks involving more fluid and less

analytical processing could be free of reactivity, as online measures are likely to promote an analytical, reflective orientation to a task.

#### *1.1.3.b. Offline measures*

Offline reports avoid the issue of reactivity by asking for information after a task is completed. A representative example of this methodology is Schmidt and Frota's (1986) diary study of Schmidt's experiences while learning Portuguese in Brazil. A notable finding was that even Schmidt, who presumably had a high degree of metalinguistic awareness, often failed to notice high-frequency grammatical structures in naturally-occurring discourse but did subsequently notice them repeatedly after learning them in class. Their study thus suggests that explicit instruction facilitates the subsequent noticing of forms within naturalistic input. In their reflections on Schmidt's in- and out-of-class learning, explicit instruction proved to be so essential to subsequent acquisition that the researchers could only find a single example of something learned that had not been taught.

Diary studies and other offline studies undoubtedly capture instances of noticing as the very ability to report on linguistic forms encountered in previous input implies a significant level of awareness. The methodological weakness of such studies derives mainly from their failure to capture the full set of instances being targeted. When noticing of an item is only fleeting, subjects may fail to report it. Moreover, subjects may, without intending to do so, creatively reconstruct the seemingly haphazard nature of online thought processes when asked to reflect on them. The unconscious tendency to "clean up" discordant features of online processes in favor of accounts that appear to make more sense as a narrative of subjective experience may lead participants to skip over features of their experience that seem (to them) to be trivial while focusing on features that they regard as significant. Learners may also be unable to adequately verbalize their underlying thought processes (Lyons, 1986), a problem for both online and offline reports. For this reason, such reports are generally appropriate only when coarse-grained measures are adequate and when samples of the *content* of noticing (versus the *process* or *scope*) constitute the research focus.

#### *1.1.3.c. Combined measures*

Some researchers have attempted to compensate for the coarse-grained nature of online and offline measures by combining them. This triangulation approach was used by Mackey (2006), who examined the effects of feedback on the noticing and learning of English question forms, plurals, and past tense. To determine participants' noticing, she used online learning journals, oral stimulated recall, questionnaires, and written/oral responses to

questions about the experimental instruction. After aggregating data from the four measures, she found that feedback promoted noticing and that this noticing, at least in the case of question forms, promoted learning.

Mackey's study has increased reliability due to her use of multiple data sources; yet as Mackey herself mentions, even this battery of measures is unlikely to capture all instances of noticing. When subjects are asked to fill out online reports, they may forget to do so if they are deeply involved in classroom activities. Stimulated recall, on the other hand, is often of little value due to time constraints: Researchers and subjects rarely have the time to laboriously work through representative segments captured on video. An even greater problem is that the above measures all fail to capture instances of detection without noticing. Mere detection must therefore be inferred, which is problematic as noticing has only been measured in an extremely coarse manner.

#### *1.1.3.d. The RK Paradigm*

In order to avoid some of the problems cited above, this study used an offline protocol based on the RK distinction. The idea of using "remember" (R) and "know" (K) states in laboratory studies was suggested by Tulving (1985) who claimed that these two states of awareness reflect auto-noetic and noetic states of consciousness, which respectively characterize episodic and semantic memory systems. In a typical test of RK states (e.g., Gardiner, Ramponi, & Richardson-Klavehn, 2003) subjects are instructed about the RK distinction in some detail prior to the test and are then asked to report on various items, to some of which they have previously been exposed. Subjects are usually asked if the word previously appeared. If they answer "yes," they are given the further option of selecting a "remember" or "know" response. Some researchers also include a "guess" response to capture instances in which participants are less confident of their response. The RK protocol, unlike other measures of noticing, thus has the ability to measure attention (in the form of "know" or "guess" responses) that falls below the threshold of awareness that is the hallmark of noticing.

Reports employing the RK distinction provide a useful means of tapping into differences in ways that information is stored, processed, and retrieved. According to Gardiner and Richardson-Klavehn (2000: 229), "remembering" involves "intensely personal experience of the past" in which we "recreate previous events and experiences with the awareness of reliving" these events mentally. "Knowing," on the other hand, is described as experiences of the past in which "we are aware of knowledge that we possess but in a more impersonal way." The "remember" versus "know" distinction thus coincides with the distinction that Tomlin and Villa (1994) make between mere *detection* and *awareness* (i.e., noticing).

The RK paradigm has not been adequately explored within SLA research. A rare exception is Peckham's (2000) study on the effect of explicit instruction on the noticing and acquisition of vocabulary and grammar. Peckham used an RK protocol to examine the noticing of Hungarian secondary school learners of English as a foreign language in instructed or non-instructed conditions. His results showed that instruction produced greater noticing than mere exposure for both grammar and vocabulary. As would be expected due to its less abstract nature, vocabulary proved easier to notice than grammar and the effects of noticing on learning were also greater.

## **I.2. Target structure**

Simard and Wong (2001) have argued that psychological models of attention, when applied to the SLA context, should account for differences among tasks, linguistic forms, and individual learners. This study has addressed the first two concerns by selecting a task (reading) and structures (word sequences) that are conducive to incidental learning through the implicit and explicit processes associated with mere detection and noticing. Confounds associated with individual differences have been accounted for through the use of a within-subjects design which contrasts the hypothesized effects within each individual participant. Lexical items have been selected over grammatical rules as they are felt to be a more realistic target of short-term instructional interventions.

## **II. EXPERIMENT 1**

To assess the effects of explicit instruction on noticing, Experiment 1 set out to determine whether metaphorical word sequences (WSs) that occurred in a reading text would be noticed (i.e., "remembered" on the RK protocol) more frequently if their variant form was previously the target of instruction.

### **II.1. Method**

#### ***II.1.1. Participants***

Thirty-six non-native speakers of English attending a university in the U.S. were recruited. Three were replaced after failing to attend the second session. The 36 participants were mostly graduate students in a master's program. They were paid \$15 for their participation.

Participants filled out a short questionnaire that asked their age, previous English study, first language, and length of residence (LOR) in an English speaking country. Their mean age was 28.5. On average, they had spent 12.9 years studying English. When asked their mother tongue, most (27) were Chinese speakers with the others reporting a wide assortment of L1s. Their mean LOR was 14.2 months but ranged from one month to 10.5 years.

### **II.1.2. Research Design**

The experiment involved a balanced random-block within-subjects design that consisted of four conditions: (1) In the instruction-plus-exposure (IE) condition, participants received explicit instruction on 11 variant forms of 11 target metaphorical WSs that occurred in the reading text. These variant forms shared one word with the target WS and, more importantly, shared the same underlying motivation. For example, the target item *nagging doubts* had a variant form: *nagging suspicion*. The semantic motivation for both forms was the fact that negative cognitive states (*doubts* and *suspensions*), like *nagging*, can recur regularly and can be annoying; (2) In the exposure-only (EO) condition, participants were exposed to 11 target WSs that occurred solely in the reading text; (3) In the instruction-only (IO) condition, subjects received instruction on 11 WSs that did not appear in the reading; (4) In the distractor (D) condition, subjects received neither instruction on nor exposure to 11 WSs. (See Appendix A). Items used for the first and second condition were counterbalanced in two blocks in order to eliminate possible differential effects stemming from differences between the items.

### **II.1.3. Procedure and Measures**

Participants received instruction on word sequences. All participants next read a story text containing 22 target WSs. They were informed prior to reading that they would be asked three general comprehension questions that they would have to answer without looking back at the text. To prevent participants from associating the tasks, they were told that the reading was a different task that simply assessed their reading comprehension. Regarding the experimental task on the following day, they were only told that they would “fill out a short questionnaire about their language background and do a few other related tasks.”

On the following day, all participants completed a 44-item RK protocol followed by a short retrospection questionnaire and then a fill-in-the-blank test. The sequencing of instruction is shown in Table 1.



Experiment Begins	One Day Later
Instruction (88 min.)	RK Protocol (20 min.)
Break (10 min.)	Retrospection (5 min.)
Reading task (8 min.)	Questionnaire (5 min.)
	Break (10 min.)
	Fill-in-Blank Test (20 min.)

Table 1. Sequencing of Measures and Treatment

#### *II.1.4.a. Explicit Instruction*

Each treatment group was taught 22 randomly sequenced WSs during the instructional phase of the experiment. Among these WSs, 11 items also appeared in the reading. The instruction included pictures and a description designed to reveal the relationship between the concepts that motivated the target metaphor (See Appendix C). Participants were given four minutes to read each paragraph and write a response

#### *II.1.4.b. Reading Task*

After a short break, participants were given eight minutes to read a fictional account of a Russian soldier's experiences during the Crimean War (see Appendix B). The text was designed to include a large number of difficult words and expressions so that the target WSs would not stand out as especially salient solely due to the difficulty that the items posed relative to the other words in the text. The text contained 22 target WSs, only half of which occurred in each group's instruction.

#### *II.1.4.c. RK Protocol*

The RK protocol included 44 items in pseudo-random order, drawn evenly from four categories as shown in Table 2.

Category	Instruction	Exposure	Group	
			I	II
IE <sup>a</sup>	+	+	1	2
EO <sup>b</sup>	-	+	2	1
IO <sup>c</sup>	+	-	3	3
D <sup>d</sup>	-	-	4	4

Table 2. RK Protocol Item Distribution

<sup>a</sup>Instruction and exposure items

<sup>b</sup>Exposure only items

<sup>c</sup>Instruction only items

<sup>d</sup>Distractor items



Participants were asked to mark a box next to YES if they believed that the WS occurred in the story text or NO if they did not think it appeared. If they chose the former response, they were asked to select REMEMBER, KNOW, or GUESS. Participants were instructed in detail regarding the distinction between each category using instructions adapted from those in Peckham (2000: 255-257). These were provided to participants in written form and were read and discussed with them immediately prior to their completion of the RK protocol.

#### *II.1.4.d. RK Retrospection*

After participants completed the RK protocol, they filled out a short retrospection questionnaire that asked them to recall a “remember,” “know,” and “guess” response and explain their reasons for selecting each of these responses.

#### *II.1.4. e. Fill-in-the-Blanks Test*

Following completion of the RK protocol and after taking a short break, participants were asked to complete a 22-item fill-in-the-blanks test requiring them to supply the original WS from the reading. The missing word from each WS was always the word common to both the WS in the instruction and the WS in the text. Participants received one point for each correct item.

## **II.2. Results and Discussion**

To check whether participants achieved adequate understanding of the story, their three general comprehension responses were scored, giving one point for responses reflecting a good understanding of events, a half point for an incomplete understanding, and zero points for misunderstandings or no response. From a total possible of three points, participants had a mean score of 2.83 ( $SD = 0.27$ ), demonstrating adequate understanding of the text. Word counts were tallied on their writing during the instruction task. On the 22 items, participants wrote a mean of 26.6 words, demonstrating adequate engagement in the task.

The RK responses were tallied. Figure 1 shows the percentage of each of the four response types on the RK protocol across the four experimental conditions.

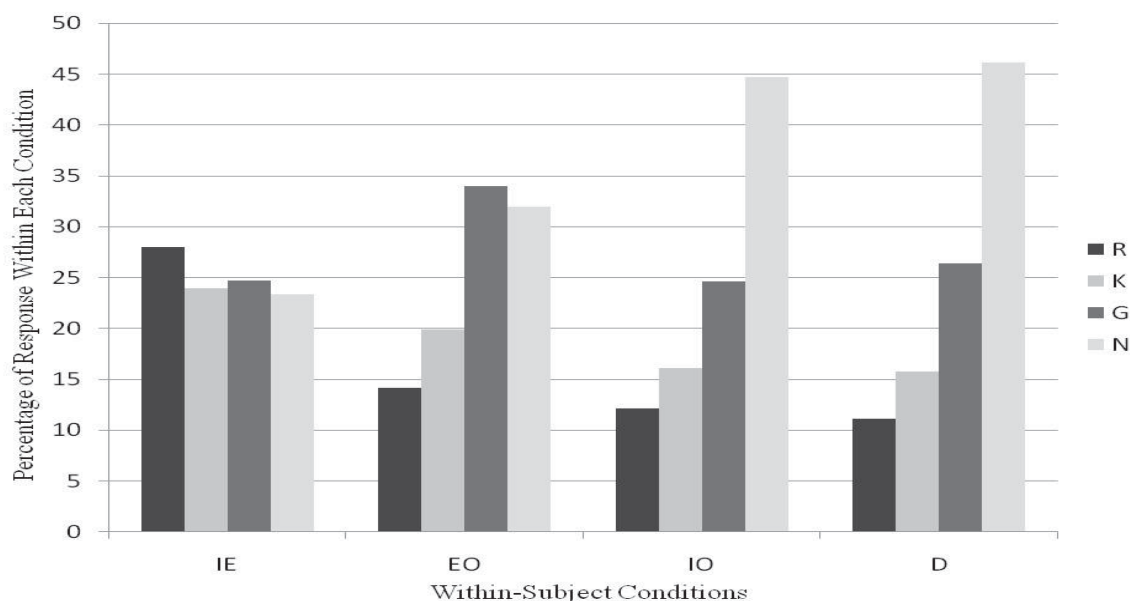


Figure 1. Percentage of “remember” (R), “know” (K), “guess” (G), and “no” (N) responses in the four experimental conditions.

As can be seen, participants gave “remember” responses to 42.1% of items from the reading (i.e., the IE and EO categories combined), with IE items accounting for roughly two-thirds of these responses. There were considerably more “guess” responses in the EO condition, suggesting that mere exposure in the reading had created a vague sense of familiarity that did not result in high levels of confidence. The IO and D results were virtually identical, an indication that instruction did not lead to more “false positives” (R, K, or G responses) in the IO condition.

A key research question in Experiment 1 was whether explicit instruction would lead participants to notice related word sequences in a subsequent reading task. A comparison was therefore conducted, examining the number of “remember” responses that each participant made for each category of item. Each category contained 11 items. Only the IE and EO items actually appeared in the reading, so participants’ choice of IO and D items represented “false positives.” The descriptive statistics for the 36 participants are shown in Table 3.

“Remember” Responses					
	<i>M</i>	%	Minimum	Maximum	<i>SD</i>
IE	2.86	28.0%	0	10	2.24
EO	1.56	14.2%	0	7	1.59
IO	1.33	12.1%	0	6	1.59
D	1.22	11.1%	0	5	1.76

Table 3. “Remember” Responses to Word Sequences That Appeared in the Four Experimental Conditions

As can be seen, participants were more than twice as likely to notice a WS after encountering a similar WS in the explicit instruction session. Moreover, “remember” responses to items that were only encountered in the reading were only marginally higher than such responses to items that did not appear (i.e., false positives). To determine whether these effects were statistically significant, a repeated measures ANOVA was conducted treating Treatment Type as a within-subjects variable with four levels: (IE) instruction and exposure, (EO) exposure only, (IO) instruction only, and (D) distractor. To test sphericity, a Mauchly’s Test was conducted. The test was significant,  $W = .61$ ,  $\chi^2(5) = 16.57$ ,  $p = .005$ , suggesting that the observed matrix does not have approximately equal variances and equal covariances. A Greenhouse-Geisser correction was therefore applied so as to raise the critical  $F$  value needed to reject the null hypothesis.

The ANOVA indicated that Treatment Type had a significant effect on incidental noticing as measured by “remember” responses on the RK measure,  $F(2.3, 79.4) = 10.47$ ,  $p < .001$ ,  $\eta^2 = .54$ . As indicated by partial eta-squared, Treatment Type was able to account for more than half of the variance within subjects. Pairwise analyses using a two-tailed test of significance at the  $p < .05$  level and a Bonferroni adjustment for multiple comparisons indicated higher scores for the IE condition compared to the EO condition,  $p = .002$ ,  $SE = .375$ , 95% CI [0.48, 2.58]; higher scores for the IE condition compared to the IO condition,  $p < .001$ ,  $SE = .262$ , 95% CI [0.91, 2.37]; and higher scores for the IE condition compared to the D condition,  $p = .003$ ,  $SE = .457$ , 95% CI [0.47, 3.03]. All other comparisons were non-significant at a  $p < .05$  level, two-tailed.

Participants’ scores on the fill-in-the-blanks test were tallied, giving full credit if the word matched that of the original reading. Responses using the incorrect word form (e.g., “fierce loyal” instead of the intended “fiercely loyal”) or incorrect spelling were counted as correct. The 36 participants were accurate on 47% of the IE items ( $M = 5.14$ ,  $SD = 2.8$ ) and on 30% of the EO items ( $M = 3.25$ ,  $SD = 2.6$ ). To statistically determine whether Treatment Type had an effect on learning as measured by the fill-in-the-blanks test, a paired-samples  $t$ -test was conducted treating Treatment Type as a within-subjects factor with two levels (IE and EO) and the fill-in-the-blanks scores as the dependent variable. As would be expected, instruction promoted learning,  $t(35) = 6.08$ ,  $p < .001$ . Among correct responses, 30.8% ( $M = 1.6$ ) of the items appearing in the IE condition and 23.1% ( $M = 0.8$ ) of the items appearing in the EO condition involved word sequences to which the participant had given a “remember” response on the RK protocol. This suggests that noticing had a slightly greater facilitative effect on the learning of IE items.

At the end of the RK protocol, participants were asked to write down their reasons for one item from each response category. The responses are given in Table 4.

Number of Responses	Reason for Response
17	Item formed a key part of the story and was thus memorable.
6	Item was previously studied.
6	Item's position in the text was recalled.
5	Item was unknown or unusual.
2	Other details of the item made it memorable.

Table 4. Experiment 1 Reasons for "Remember" Responses

Peckham's (2000: 92) study reported as key reasons given for a "remember" response items' having been "studied before" or having been "unknown or unusual" or the participant's ability to "remember seeing/visualizing it in the text." While many of the responses were similar, the participants in this experiment seemed to rely more on the word sequence's relationship within the story. It should be noted that only half the items had been previously studied, so the six "studied before" responses are roughly equivalent to 12 responses in the other categories.

When asked to describe a "know" response, 20 participants made a statement to the effect that they "just knew" that the word sequence was in the story but did not recall the location or details. Other responses were diverse.

When describing "guess" responses, 22 participants made a statement related to the word sequence's plausibility within the story. About a third of the responses contained language expressing lack of confidence in their response and three participants specifically mentioned an inability to determine the source of the memory (e.g., the study materials or the reading text).

Experiment 1 confirmed that explicit instruction facilitates incidental noticing. Participants' RK retrospections suggest that the RK protocol may be reliably used with this population to measure incidental noticing.

### III. EXPERIMENT 2

If explicit instruction triggers noticing, an important question is whether explicit instruction that is thought to produce deeper levels of processing also promotes more incidents of noticing. Specifically, would instruction using self-referential writing prompts result in more elaborative encodings that would, in turn, result in more episodic memory traces?

A number of studies, largely inspired on Craik and Tulving's (1975) depth-of-processing model, have found better recall for self-referential items (Klein & Loftus, 1988; Symons & Johnson, 1997). This effect has been explained in terms of either greater

elaboration or in terms of more detailed structural organization of memory schemas related to the self. In SLA pedagogy, the effect has led to a preference for language learning tasks that promote personal involvement. Experiment 2 therefore attempted to determine whether self-referential materials would lead to more incidents of incidental noticing on a subsequent reading task.

### **III.1. Method**

Except for a few minor modifications, Experiment 2 largely replicated the conditions of Experiment 1.

#### ***II.1.1. Participants***

For the second experiment, 24 college students who had not participated in Experiment 1 were recruited from a large U.S. university. Two participants were excluded when they failed to attend the second session and two were excluded based on reasons they provided for their RK responses that indicated that they had misunderstood the RK instructions. These four were replaced. The resulting 24 participants consisted of 12 Chinese L1 speakers, 11 Korean L1 speakers, and one Farsi L1 speaker. Participants had a mean age of 29. On average, they had spent 12 years studying English and had spent 4.5 years in an English speaking country. Nearly all were matriculated college students. All were paid \$20 for their participation.

#### ***II.1.2. Research Design***

Experiment 2 split the IE category of Experiment 1 into two categories: (1) a deep processing condition and (2) a shallow processing condition. As in the first experiment, the IE and EO word sets were counterbalanced. To elicit deep processing during the instructional phase, half of the IE items (the “deep” items) were followed by questions that encouraged participants to make personal connections between the WS’s meaning and their own experiences, values, aspirations, and so on. The other half of the instructional items (the “shallow” items) were followed by questions that were, as far as possible, similar, but which did not encourage self-reference. Both deep and shallow versions of each item were created and these were also counterbalanced across sets. IO items were also split between deep and shallow versions that were similarly counterbalanced.

#### ***II.1.3. Measures and Procedures***

The measure and procedures were identical to those of Experiment 1 except that an additional word sequence was added to the four categories (i.e., IE, EO, IO, and D) and the word lists of

the IE and EO categories were shuffled and reconstituted. Participants were given three minutes to complete both the reading and writing task for each target item. In addition, the RK protocol and fill-in-the-blanks test were expanded to reflect the addition of four items.

### III.2. Results and Discussion

To confirm participants' engagement in the reading task, their responses to the general reading comprehension questions were scored applying the same three-point system used in Experiment 1. Participants' mean score was 2.83 ( $SD = 0.20$ ), indicating adequate understanding and engagement in the task.

To determine whether the participants ( $N = 24$ ) produced more words on deep versus shallow items during the instructional writing task, word counts for both conditions were compared. Participants' mean word count for deep items was 436.2 ( $SD = 117.9$ ), or 18.2 words per item, compared to a mean word count of 408.7 ( $SD = 105.4$ ), or 17.0 words per item, for shallow items. Treating Question Type as a within-subjects factor with two levels and Word Count as the dependent variable, a two-tailed paired samples  $t$ -test using an alpha level of  $p < .05$  indicated that participants wrote significantly more words in the deep condition,  $t(23) = 3.29$ ,  $p = .003$ , 95% CI [10.2, 44.8].

A further question was whether self-referential writing prompts would lead to more noticing. To test this, the number of RK "remember" responses was compared for items in each of the four within-subjects conditions. The results are provided in Table 5.

		"Remember" Responses				
		<i>M</i>	%	Min.	Max.	<i>SD</i>
Total	48	6.5	14%	0	17	4.0
IE Total	12	3.7	31%	0	8	2.4
IE SR <sup>a</sup>	6	1.7	28%	0	5	1.5
IE Non-SR <sup>b</sup>	6	2.0	33%	0	4	1.5
IO Total	12	0.7	6%	0	5	1.2
IO SR <sup>a</sup>	6	0.2	3%	0	3	0.7
IO Non-SR <sup>b</sup>	6	0.5	8%	0	3	0.8
EO Total	12	1.8	15%	0	8	1.9
D Total	12	0.4	3%	0	2	0.6

<sup>a</sup>Self-referential writing prompts

<sup>b</sup>Non-self-referential writing prompts

Table 5. "Remember" Responses to Word Sequences That Appeared in the Four Experimental Conditions in Experiment 2

As can be seen in Table 5, participants reported "remembering" 14% of all items. Contrary to expectations, the IE shallow items were remembered slightly more often than the IE deep

items, but this small gap disappears if “false positives” (falsely “remembered” items that were not in the reading) from the deep and shallow IO items are subtracted from the respective IE totals. This would leave a mean of 1.5 IE deep items ( $SD = 0.7$ ) and 1.5 IE shallow items ( $SD = 1.7$ ). The table’s results have been graphically depicted in Figure 2.

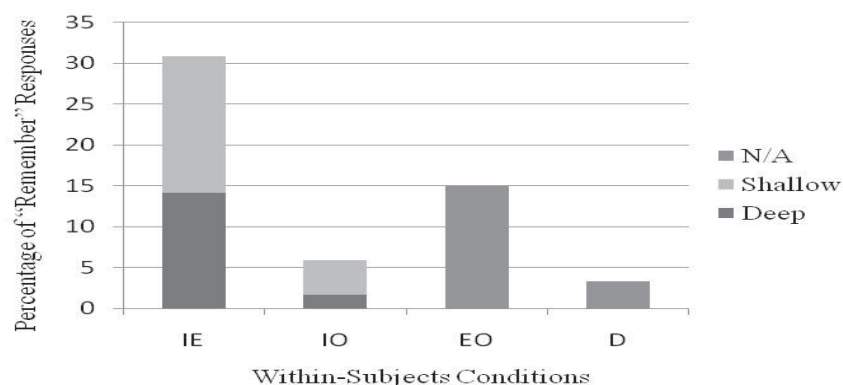


Figure 2. Percentage of “remember” responses for each of the four within-subject conditions in Experiment 2.

The figure shows the four main conditions as columns. The two columns on the left have been further divided, showing the difference between items studied in the shallow (light shading) versus deep (dark shading) condition. As can be seen, participants tended to notice IE WSs about twice as often as they noticed EO WSs. This is consistent with the results of Experiment 1. As would be expected, “false positives” (false remembering of words in the IO and D conditions) tended to occur more often if the WS occurred in the IO condition. It is interesting to note that more of the IO false positives occurred in the shallow condition as opposed to the deep condition.

Statistical tests were conducted to determine whether Question Type was related to “remember” responses on the RK protocol. The greater number of “remember” reports for studied items (i.e., those in the IE condition) relative to unstudied items (i.e., those in the EO condition) was significant at  $p < .05$  two-tailed,  $t(23) = 4.148$ ,  $p < .001$ , 95% CI [0.96, 2.87]. This confirms the main finding of Experiment 1 (i.e., explicit instruction’s promotion of noticing).

In the IE condition, participants ( $N = 24$ ) gave “remember” responses to 16.0% of deep items ( $M = 1.92$ ,  $SD = 1.64$ ) and to 20.5% of shallow items ( $M = 2.46$ ,  $SD = 1.67$ ), but these differences failed to reach significance at  $p < .05$  on a two-tailed paired-sample  $t$ -test,  $t(23) = -1.406$ ,  $p = .173$ , 95% CI [-1.34, 0.26]. Incidents of “false positives” occurred with 3.5% ( $M = 0.21$ ,  $SD = 0.66$ ) of the IO deep items compared to 7.6% ( $M = 0.46$ ,  $SD = 0.17$ ) of the IO



shallow items, but the difference was nonsignificant at  $p < .05$  two-tailed,  $t(23) = -1.366$ ,  $p = .185$ , 95% CI [-0.63, 0.13].

On the RK retrospection, responses were similar to those of Experiment 1 with nine participants responding that they selected a “remember” response because it formed a key part of the story and five saying that they chose “remember” because the item had been studied before.

On the fill-in-the-blank test, participants were accurate on marginally more IE items (42.0%,  $M = 5.0$ ,  $SD = 2.5$ ) than EO items (40.3%,  $M = 4.8$ ,  $SD = 2.6$ ), but the difference was nonsignificant at  $p < .05$  on a two-tailed paired samples  $t$ -test,  $t(23) = 0.385$ ,  $p = .704$ , 95% CI [-0.91, 1.33]. Participants were more accurate on the fill-in-the-blanks items which had been studied in the IE deep condition (46.5%,  $M = 2.79$ ,  $SD = 0.4$ ) compared to those in the IE shallow condition (37.5%,  $M = 2.25$ ,  $SD = 1.3$ ), but the difference was nonsignificant at  $p < .05$  on a two-tailed paired samples  $t$ -test,  $t(23) = 1.423$ ,  $p = .168$ , 95% CI [-0.25, 1.33].

Experiment 2 replicated Experiment 1 results, showing that explicit instruction facilitates incidental noticing, but it failed to show that self-referential prompts lead to more noticing or more learning. The 9% higher scores on the fill-in-the-blanks test for deep items suggest the possibility of an effect that was not large enough to detect due to the low sensitivity of the statistical test, which involved only six items in each category of the relevant comparison.

#### IV. CONCLUSION

The two experiments demonstrated that explicit instruction, in addition to any immediate effects, may have important lingering effects by altering learners' sensitivity to important semantic mappings within the input. The findings thus suggest a synergy between explicit instruction and subsequent communicative practice that involves the same (or similar) target forms. If these lingering effects are confirmed in future research, language instructors may feel more justified when using explicit instruction to target moderately difficult linguistic patterns even when the instruction does not lead to immediate acquisition.

The findings for self-referential writing prompts were mixed. It could be that self-referential questions do, in fact, produce more semantic elaboration but that this elaboration has little impact for learners who are encountering a WS for the first time. This interpretation is suggested by a body of research (for a review, see Barcroft, 2002) showing that semantic elaboration facilitates learners' memory of known words but does not facilitate memory of unknown words. Although the individual words in the WSs used in the two experiments were

likely to be familiar to most participants, the meanings of the WSs themselves were probably unknown. On the other hand, Experiment 2 suggests that self-referential prompts may be of general benefit to learners by encouraging more linguistic production.

The generalizability of the findings of both experiments is limited by several factors. First, the participants nearly all spoke non-Indo-European languages that had few English cognates. Explicit instruction focusing on metaphorical WSs may have different effects on speakers of Indo-European languages, who are more likely to be affected by L1 transfer due to perceived or actual similarities between their L1s and English. More research also needs to be conducted on other linguistic forms and tasks (e.g., listening tasks).

Turning to methodology, the study has demonstrated the feasibility of using the RK paradigm in SLA research. Cognitive states are intrinsically fleeting and difficult to measure; no behavioral measure of noticing, including the RK protocol, is likely to be process-pure. Even so, the RK paradigm should be explored further as it opens the possibility of fine-grained explorations of degrees of attention that are inaccessible to other methodologies currently in use.

## REFERENCES

- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, 52(2), 323-363.
- Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 227-248). Amsterdam: John Benjamins.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268-294.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.
- Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). Cambridge, UK: Cambridge University Press.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2003). Experiences of remembering, knowing, and guessing. In B. J. Baars, W. P. Banks & J. B. Newman (Eds.), *Essential sources in the scientific study of consciousness* (pp. 697-720). Cambridge, MA: MIT Press.
- Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244). Oxford, UK: Oxford University Press.
- Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Cripe & A. P. R. Howatt (Eds.), *Interlanguage: Studies in honor of Pit Corder* (pp. 291-311). Edinburgh, Scotland: Edinburgh University Press.
- Kim, H. (1995). Intake from the speech stream: Speech elements that L2 learners attend to. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 65-83). Honolulu, HI: University of Hawai'i.
- Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology*, 55(1), 5-11.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon.

- Leow, R. P. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's (1994) fine-grained analysis of attention. *Applied Psycholinguistics*, 19(1), 133-159.
- Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, 84(3), 496-509.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26(1), 35-57.
- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. De Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in crosscultural perspective* (pp. 39-52). Amsterdam, The Netherlands: John Benjamins.
- Lyons, W. (1986). *The disappearance of introspection*. Cambridge, MA: MIT Press.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, 27(3), 405-430.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- Peckham, D. (2000). *Attention and consciousness in second language acquisition: An investigation into the effects of instruction on noticing*. Unpublished Doctoral Dissertation, University of Pittsburgh.
- Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, 33(4), 465-497.
- Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition*, 21(4), 511-556.
- Schmidt, R. (1983). Interaction, acculturation and the acquisition of communicative competence. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11-26.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.
- Schmidt, R., & Frota, S. N. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.
- Simard, D., & Wong, W. (2001). Alertness, orientation, and detection: The conceptualization of attentional functions in SLA. *Studies in Second Language Acquisition*, 23(1), 103-124.
- Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, 121(3), 371-394.
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183-203.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1-12.
- VanPatten, B. (1996). *Input processing and grammar instruction*. Norwood, NJ: Ablex.

## Appendix A

### Target Word Sequences (WSs) and Variant Forms

Reading Text/RK	Instruction	Set
air of authority	air of elegance	1
biting wind	biting cold	1
borrowed time	borrowed moments	1
bout of depression	bout of fever	1
brave souls	free souls	1
burning ambition cold	burning desire	1
gaze	icy gaze	1
deeply troubled harsh	deeply agitated	1
reality	harsh facts	1
tortured logic	tortured reasoning	1
waves of attacks blind	waves of people	1
faith	blind obedience	2
blissfully unaware	blissfully ignorant	2
crushing defeat empty	crushing blow	2
promises fallen	empty words	2
soldiers fiercely loyal	fallen comrades	2
high ideals	fiercely independent	2
nagging doubts	high aspirations	2
planted firmly	nagging suspicion	2
read his mood	planted solidly	2
sobering reminder	read someone's face	2
cooked up an excuse	sobering news cooked	2
fragile truce	up a scheme fragile	3
fueled by passion hail	agreement fueled by	3
of bullets	greed	3
hazy recollections	hail of arrows	3
heap praise (on)	hazy memories	3
landed a promotion	heap scorn (on) landed	3
painfully honest sliver	a job	3
of hope surgical	brutally honest	3
bombing tackle the	sliver of truth surgical	3
issue	strikes	3
baby steps	tackle the problem	3
bragging rights display		4
tact		4
dreary tale grudgingly		4
tolerate hardened		4
warriors locked in		4
combat sink into		4
madness sounded the		4
alarm starry-eyed		4
theater of war		4
		4
		4

\* Experiment 2 shuffled the categories and then added “heavy losses” (heavy costs) to Set 1, “bitter fight” (bitter dispute) to Set 2, “crooked man” (crooked leaders) to Set 3, and “stormed the barricades” to Set 4.

## Appendix B

### Reading Used on First Day of Both Experiments

#### A Soldier's Account of the Crimean War

*The following paragraph is a fictional account of a soldier fighting for the Russian forces during the Crimean War during which Russia fought against Britain, France, and the Ottoman Empire. Read the passage. When you are finished, you will be asked three general questions about the reading.*

We were under the command of Nikolai Muravyov, a brilliant strategist who had a certain air of authority that inspired the utmost confidence in his troops. Muravyov was fiercely loyal to our nation's cause, with a burning ambition to crush the French and English devils and re-establish Russian protection over the holy sites of Constantinople. We all knew that Russia had been the protector of Orthodox Christians in the Holy Land from early on, based on a 1757 treatise. From the very beginning of the campaign, the general was wildly popular with the troops, who remained blissfully unaware of the harsh reality of war. Fresh from our training, we all had high ideals about our duty to God and country. In a sense, we were perfect soldiers—willing to die and willing to disregard the negative reports of injured soldiers returning from the front.

In our first battle, we fought against a Turkish force led by William Fenwick Williams, a British commissioner, who had recently been made general of the combined British and Turkish forces. Our direct assault on a well-prepared enemy proved costly, as the British, with superior weaponry and excellent battle tactics, fought off waves of attacks before going on the offensive. In the end, our unit experienced a crushing defeat in which we lost many of our best men and horses. My cousin Ivan became deeply troubled after watching the slow and painful death of his horse, a beautiful stallion that had been a gift from the general himself.

As our unit gathered into formation after the battle, the youthful drummer who had led us into battle on the previous morning was conspicuously absent, a sobering reminder of our unit's heavy losses. I have always considered myself as valiant and as patriotic as any of my Russian countrymen, but I must confess that I was, after watching the needless loss of so many brave souls, becoming bothered by nagging doubts about the war. I couldn't help feeling that I was living on borrowed time, that I would soon become just another nameless fighter in another nameless grave. That night I sat with my cousin Ivan amidst my injured countrymen. Beyond our campfires, I could see packs of dogs tearing at the limbs of the fallen soldiers, who were scattered here and there across the battlefield. In the officers' tent in the distance, I overheard men engaged in a bitter dispute over the battle plan as flickering light from a lamp danced around their tent. Ivan sat rigidly as he coldly gazed at the scene, seemingly oblivious to all the suffering around him. I knew he was affected—any man would be—but I couldn't quite read his mood. He seemed strangely aloof.

In the following weeks, we brought our forces together and prepared for the long battle to come. As summer turned to autumn, I stood one evening with Ivan in the biting wind coming off the Black Sea. I had known through my nightly chats with my cousin that he was growing impatient with the war, but this evening his expression was oddly tranquil. He appeared in the soft moonlight as if he were half-man and half-ghost, with one foot already planted firmly in the next world. In the ensuing weeks, as we slept each night in our tattered tent, I would watch him toss and turn as if fighting off armies of inner demons.

One morning, after a particularly serious bout of depression, Ivan walked up to three of us while we were preparing our meal. He had taken a bath and his uniform was creased and ironed. In an overly thoughtful manner, he gave us his remaining ration of tea—a prized commodity on the battlefield. His generosity might have passed unnoticed had he not then turned and handed me a sealed letter addressed to his wife and children. An hour later as we marched back out to the front lines, I heard the lone crack of a gun going off in the distance. Some of our men ran towards Ivan's tent, but I didn't even turn around; I already knew what had happened. Ivan, who had emerged unscathed from every battle thus far, had suffered defeat in the most difficult battle of all—the fight to remain sane in an

insane world where men kill one another based on the empty promises and tortured logic of their leaders and blind faith in the righteousness of their cause.

1. Ivan must have witnessed many deaths and tragedies during the war. What particular event do you think led him to commit suicide?
2. Do you think the person narrating the story knew that Ivan would commit suicide? If so, why didn't he do more to prevent Ivan from killing himself?
3. When the person narrating the story returns home, how will he describe the war? How will the war have affected his ideas about patriotism and honor?

## **Appendix C**

### Sample Treatment Item

#### **Fueled by Greed**

Cars are typically fueled by gas or other fuels. These fuels, when put into the car, make the car move. People can also move based on their inner emotions, which make people act in certain ways. The rich man on the right is **fueled by greed**. He wants to make more money, and this greed serves as the "fuel" that keeps him going throughout the day as he pursues ways to make ever greater profits.

WRITING TASK: Some people would say that modern economies are **fueled by greed** and that greed is not necessarily a bad thing since it encourages competition. What do think about this?

[In the instructional materials, this item was flanked by a picture of a gasoline nozzle and another picture of a greedy businessman clutching dollar bills.]

