SSLA, **24**, 541–577. Printed in the United States of America.
DOI: 10.1017.S0272263102004023

OUTPUT, INPUT ENHANCEMENT, AND THE NOTICING HYPOTHESIS

An Experimental Study on ESL Relativization

Shinichi Izumi
Sophia University

This study investigates the potentially facilitative effects of internal and external attention-drawing devices—output and visual input enhancement—on the acquisition of English relativization by adult English as a second language (ESL) learners. Specifically, the study addresses: (a) whether the act of producing output promotes noticing of formal elements in the target language (TL) input and affects subsequent learning of the form; and (b) whether such output-induced noticing and learning, if any, would be the same as that effected by visual input enhancement designed to draw learners' attention to problematic form features in the input. These questions were examined in a controlled experimental study in which the requirements of output and exposure to enhanced input were systematically varied. A computer-assisted reconstruction and reading task was used as the vehicle of presentation of the target input materials. The major findings are: (a) Those engaged in output-input activities outperformed those exposed to the same input for the sole purpose of comprehension in learning gains; (b) those who received visual

An earlier version of this paper was presented at the annual meeting of the American Association for Applied Linguistics held in St. Louis, MO, in February of 2001. This paper is based on part of the author's doctoral dissertation, which was completed at Georgetown University. I am thankful to Catherine Doughty for her guidance throughout the entire process of my research project. My thanks also go to Jeff Connor-Linton and Cristina Sanz for their critical comments and encouragement. Thanks are also due to a number of people who were directly or indirectly involved in this research project, including the directors, teachers, and students of the participating ESL programs as well as people who were kindly involved in the creation, execution, and assessment of the tests and tasks used in the study. Finally, I wish to thank the anonymous *SSLA* reviewers for their helpful comments and insights.

Address correspondence to: Shinichi Izumi, Department of English Language and Studies, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554, Japan; e-mail: s-izumi@hoffman.cc.sophia.ac.jp.

© 2002 Cambridge University Press 0272-2631/02 $9.50

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

input enhancement failed to show measurable gains in learning, despite the documented positive impact of enhancement on the noticing of the target form items in the input; and (c) in view of the above, no support was found for the hypothesis that the effect of input enhancement was comparable to that of output. The subsequent discussion centers on reexamining the construct of noticing and argues for the need to consider levels and types of processing in order to account for how sensory detection can lead to learning.

For many second language (L2) learners and teachers, producing language (i.e., output) is generally considered to constitute an important part of L2 learning. However, precisely how beneficial it is to produce language is often not so clear. If the question is considered in terms of the contribution of output to the development of L2 knowledge as opposed to its utility in increasing fluency, the answer may become even more vague and speculative. Moreover, if output has any positive effect on learning, we may wish to ask whether it is unique to output or if essentially the same effect can be obtained by some external manipulation of input. These are the issues that are addressed in the study reported on in this article.

ATTENTION AND NOTICING IN SLA

The global consensus that has emerged from decades of research in SLA is that input plays a crucial role in driving learners' acquisition of an L2. Current SLA research, however, goes beyond general interest in the need for comprehensible input (Krashen, 1985), which is considered necessary but insufficient (see Ellis, 1994; Larsen-Freeman & Long, 1991; and Long, 1996, for reviews). Instead, it seeks to obtain a more precise understanding of how learners process, or interact with, input to develop their interlanguage (IL) competence. Given that not all of the input that learners are exposed to is utilized as intake for learning, recent research in cognitive psychology and SLA has examined the role of attention in mediating input and learning. A general finding of such research indicates that attention is necessary for learning to take place (see Robinson, 1995; Schmidt, 1990, 1995, 2001; and Tomlin & Villa, 1994, for reviews; see also Schachter, 1998; and Truscott, 1998, for skeptical views). Simply stated, "people learn about the things that they attend to and do not learn much about the things they do not attend to" (Schmidt, 2001, p. 30).

This general agreement on the importance of attention notwithstanding, disagreement exists as to the amount and type of attention needed for learning. Three positions have been put forth in the SLA literature. First, the Noticing Hypothesis (Schmidt, 1990, 1995, 2001; Schmidt & Frota, 1986) claims that "intake is that part of the input that the learner notices" (1990, p. 139). Noticing, Schmidt argued, crucially requires focal attention and awareness on the

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

part of the learner. A second position, proposed by Tomlin and Villa (1994), claims that, of three interrelated processes of attention—alertness, orientation, and detection—only detection, which does not require conscious awareness, is crucial for learning; the other two processes may help to increase the chance of detection and, thus, learning. Third, Robinson (1995) defined Schmidt's noticing as what is both detected and then further activated as a result of the allocation of attentional resources from a central executive. He argued that different task demands stimulate different types of further cognitive processing.

Pedagogical Attempts to Promote Learners' Noticing of Form

Taking the central role of attention in learning as a starting point of investigation, recent SLA research has begun to explore whether and how the learners' attentional processes may be influenced for the sake of their greater IL development. Such consideration is indeed at the core of influential pedagogic proposals known as consciousness-raising or input enhancement (Rutherford & Sharwood Smith, 1985; Sharwood Smith, 1993) and focus on form (Doughty, 2001; Doughty & Williams, 1998; Long, 1991; Long & Robinson, 1998).

Two specific pedagogical approaches to draw the learner's attention to form have received considerable attention in recent SLA research. One is known as visual (textual or typographical) input enhancement, and the other is learners' output. These approaches share a basic characteristic—namely, an attempt to direct the learner's otherwise elusive attention to problematic aspects in the input to promote their acquisition. They differ, however, in how this is achieved. Whereas attention in the case of visual input enhancement is induced by external means (i.e., by highlighting selected input forms), attention in output arises internally through production processes, in that learners themselves decide what they find problematic in their production and what they pay attention to in the input (although external manipulation of task variables may intervene in this process). In other words, it may be argued that visual input enhancement is an external attention-drawing technique, whereas output is an internal attention-drawing device. Previous studies on these two approaches will be examined in turn.

Visual Input Enhancement: Previous Studies and Remaining Issues

Visual input enhancement is an implicit and unobtrusive means to draw the learners' attention to form contained in the written input (Doughty & Williams, 1998). The basic method of the enhancement is simply increasing the perceptual salience of the target form via combinations of various formatting techniques (e.g., bolding, capitalizing, or underlining), which may sometimes be accompanied by an explicit mention to the learners to attend to the highlighted form. With a particular form chosen for the target, the enhancement

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

embedded in the overall reading lesson aims to achieve the integration of attention to form and attention to meaning.

Previous studies on the effects of visual input enhancement—both those that used short-term treatment with rather limited exposure to the input (Alanen, 1995; Jourdenais, Ota, Stauffer, Boyson, & Doughty, 1995; Leow, 1997; Robinson, 1997; Shook, 1994; Williams, 1999) and those that used longer-term treatment with a greater amount of input exposure (Doughty, 1988, 1991; Jourdenais, 1998; White, 1998)—produced quite mixed results. Four of these nine studies (Doughty; Jourdenais et al.; Shook; and Williams) yielded positive findings for the facilitative effect of the enhancement, whereas three studies (Alanen; Robinson; and White) showed only limited effects and two (Leow; Jourdenais) found no significant effect at all. Although various differences in these studies make direct comparison among them difficult, an examination of several factors—of which I focus on three—is instrumental in identifying the directions for future research on the issue (for a detailed review of these studies, see Izumi, 2000).

First, although clearly a desirable design feature, not all studies incorporated the noticing assessment component in their research designs. Studies that did use some kind of noticing measure provided interesting evidence as to the degree and nature of noticing. For example, Alanen's (1995) study suggested that noticing seems to be an important factor in accounting for subsequent learning, but the cross-comparison of the noticing results and learning outcomes suggests that noticing seemed to be induced by a variety of factors, not the least of which was input enhancement. White (1998), on the other hand, reported that many learners noticed the forms but were not sure of their relevance or importance, which arguably accounted for the limited improvement by the enhancement group in her study. Given the alleged importance of attention in learning that is presupposed by most studies, and given the uncertainty regarding the link between input enhancement, noticing, and learning, the employment of a noticing measure, or better yet, multiple measures of noticing, seems critical in future studies.

Second, previous research has raised questions about whether input enhancement alone can really induce desired learning effects as intended by the researcher. In this vein, Williams's (1999) study, which showed the benefits of visual enhancement, used a form-focused verbatim recall task in conjunction with visual input enhancement. Such tasks may have served as a focused output task, inducing greater noticing of the form than may have been possible by visual enhancement alone. Likewise, Doughty (1988, 1991) found positive effects of the meaning-oriented treatment that involved not only visual enhancement but also various forms of comprehension assistance given for each single sentence. As such, her meaning-oriented treatment may be construed to be more explicit, elaborated, and focused than a simple enhancement condition. Thus, it seems that further investigation into the effects of a combination of instructional techniques may be worthwhile.

Finally, for any research on the effects of intervention treatment on L2

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

learning, it is useful to examine a priori the learner's level of proficiency, or more specifically the developmental level vis-à-vis the target form, because it is quite possible that the treatment effect is constrained by the learner's developmental readiness (Pienemann, 1984, 1998). Among the nine studies reviewed here, only Doughty (1988, 1991; and perhaps White, 1998) assessed the learnability of the target form for the selected group of learners prior to the start of the treatment. A general lack of learnability considerations in many of the studies is partly due perhaps to the inherent difficulty involved in determining the learnability of forms (but see Pienemann's, 1998, processability theory) and partly due to the debate surrounding how rigidly the learnability notion can and should be applied in SLA classroom studies and pedagogical practices (see, e.g., Spada & Lightbown, 1999). Such debate notwithstanding, more studies taking the learnability notion into account would be desirable if we are to be sure that the learners are indeed psycholinguistically capable of learning the targeted form.

The present study takes these considerations into account and investigates the effects of visual input enhancement on the noticing and acquisition of a grammatical form by adult L2 learners. Another independent variable to be investigated is learner output. The inclusion of these two variables permits the investigation of how they may interact or contrast in promoting SLA.

The Role of Output in SLA

The currently popular view of output posits that it constitutes not just the product of acquisition or the means by which to practice one's language for greater fluency but also a potentially important causal factor in the acquisition process. In general, the importance of output in learning may be construed in terms of the learners' active deployment of their cognitive resources. In other words, it is posited that the output requirement presents learners with unique opportunities to process language that may not be decisively necessary for simple comprehension. In proposing the Output Hypothesis, Swain (1985) argued that producing the target language (TL) may serve as "the trigger that forces the learner to pay attention to the means of expression needed in order to successfully convey his or her own intended meaning" (p. 249). In psycholinguistic terms, it may be assumed that grammatical encoding and monitoring mechanisms (see Levelt, 1989) play particularly important roles for learning purposes by functioning as an internal priming device for grammatical consciousness-raising for the language learner (see Izumi, 2000, for a detailed discussion of the psycholinguistic rationale of the Output Hypothesis).

Of the four functions of output specified in the current version of the Output Hypothesis (Swain, 1993, 1995, 1998), the present study focuses on its noticing function. The noticing function of output posits that learners may notice the gap in their IL knowledge in an attempt to produce the TL, which then prompts them to solve their linguistic deficiency in ways that are appropriate

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

in a given context. For example, if learners are left on their own to solve the immediate production difficulties, they may engage in various thought processes that can consolidate existing knowledge or possibly generate some new knowledge on the basis of their current knowledge (Swain & Lapkin, 1995). On the other hand, if relevant input is immediately available, the heightened sense of problematicity during production may cause the learners to process subsequent input with more focused attention; they may try to examine closely how the TL expresses the intention that they just had difficulty in expressing on their own. In either case, learning is believed to be enhanced through the act of producing language, which, by its mechanisms, increases the likelihood that learners become sensitive to what they can and cannot say in the TL, which leads to the reappraisal of their IL capabilities.

Previous empirical studies on the noticing function of output have produced mixed findings. In a series of studies (Izumi & Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Fearnow, 1999), my colleagues and I investigated whether output would indeed alter the learners' subsequent input processing and promote their IL development. Focusing on the English past hypothetical conditional, these studies compared a group that was given output opportunities and subsequent exposure to relevant input and a group that received the same input for the sole purpose of comprehension. This basic format of the treatment was instantiated in two types of tasks—a text-reconstruction task and a guided essay-writing task—that were delivered in reverse orders in the two studies. The results of both studies indicated a significant improvement on the form only after the second phase of the treatment (i.e., both types of tasks), which suggested the importance of extended opportunities to produce output and receive input in effecting substantial learning.

In terms of task effects, both studies found that the essay-writing task was more susceptible to individual variation than was the text-reconstruction task. It seems that the greater freedom in production in the essay-writing task makes the IL output–TL input comparison difficult vis-à-vis the target grammatical form, leading different learners to attend to vastly different aspects of the input. On the other hand, the reconstruction task may have an advantage in promoting noticing the gap when a specific form is targeted, as these tasks maximize the similarities between the learner's production and the TL model. In general, however, output opportunities in these studies had variable effects on noticing and learning of the form for different learners, which resulted in blurring the overall between-group advantage of the output group. One proposed explanation for this result is that presenting a text-length passage to read all at once may have been exceedingly taxing on the attentional capacities of some subjects and made attending to form difficult. Another proposed explanation, which finds some support in the IL analysis by Izumi and Bigelow (2000), is that the target form in these studies may not be particularly noticeable through the process of output followed by relevant input. It was suggested that a combination of formal complexity and functional expendability of the form might have contributed to this result.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Four implications may be drawn from these studies for future research. First, it would be important to provide learners with extended opportunities to produce output and receive relevant input to ensure maximal benefit from the output-input treatment. Second, for output to promote noticing and learning of a specific form, relevant task characteristics need to be considered carefully. Also of interest in this regard is how output may be combined with other focus-on-form techniques to promote greater learning. Third, care needs to be taken to make sure that the learners' processing capacity is not overloaded during output and input processing to allow for adequate allocation of attentional resources to forms. Finally, the effects of output need to be investigated with other linguistic forms. The present study takes these points into account and aims to further our understanding of the input-output relationships. Specifically, building on earlier studies, the present study examines whether output and (visual) input enhancement, in isolation or in combination, promote noticing and learning of an L2 grammatical form.

STUDY METHODOLOGY

Target Form

The target form in this study, English relative clauses (RCs), was chosen for several reasons. Most generally, the acquisition of RCs in English has been studied extensively by a number of researchers working on both first language (L1) acquisition and SLA (see Doughty, 1988, 1991, and Izumi, 2000, for reviews of relevant studies). These studies provide a rich source of information as to the structures of English RCs, the natural developmental sequence of acquisition of these forms, the potential facilitative role of instructional intervention on them, and processing problems that both L1 and L2 learners face in producing and comprehending them.

More specifically, the typological markedness relationship that holds among different relativization types and the putative existence of natural developmental sequences of RC acquisition generate three implications that are important for determining the sequencing, timing, and assessing of the instructional treatment. For sequencing, knowledge of the developmental sequences and the markedness relationship allows for an informed decision on which structure, among many relativization types available in English, should be targeted for instruction. For timing, knowledge of the developmental sequences enables us to give careful consideration to the learnability of the form for particular groups of learners. For assessment, the markedness relationship informs us of how to assess the extent of learning caused by the instructional intervention (i.e., the degree of the generalizability of the acquisition of the targeted form to other related structures in the implicational relationship). As in Doughty's (1988, 1991) studies, the instruction of the present study focused on the object-of-preposition (OPREP) type of RCs.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Table 1. Four treatment groups included in the study

	Output	
Input enhancement	Required	Not required
Enhanced	+O+IE	–O+IE
Unenhanced	+O–IE	–O–IE

Research Design

The study was a controlled experimental study with a pretest–posttest design, involving four treatment groups and one control group (see Table 1). The treatment groups differed with respect to output requirements (notated as ±O) and exposure to enhanced input (notated as ±IE) such that: The +O–IE group was required to produce output and was exposed to regular, unenhanced input; the +O+IE group was required to produce output and received enhanced input; the –O+IE group received enhanced input without output; and –O–IE group received unenhanced input without any output requirement. The control group participated only in the pre- and posttests.

Dependent variables of the study included the noticing measures, which were used to address the extent of noticing induced by output and input enhancement, and acquisition measures, which were used to address the extent of learning brought about by the respective treatments. As will be subsequently described, two types of noticing measures were used: note scores derived from notetaking done by the subjects during the input exposure phases of the treatment, and immediate uptake of the form demonstrated in the subjects' production during the output phases of the treatment.[1] Thus, the factors included in the study were: (a) two between-groups factors—the output condition with two levels (+O and –O) and the input enhancement condition with two levels (+IE and –IE); (b) time as a within-group factor with two levels—a pretest and a posttest; and (c) two dependent measures—noticing and acquisition.

Research Hypotheses

The research hypotheses were based on the consideration of the results of the prior theoretical and empirical research previously discussed. Hypotheses 1 and 2 are concerned with the noticing issue in relation to output and input enhancement, respectively. Hypothesis 3 was formulated on the basis of the assumption that the learners' noticing would be greater if the two attention-drawing techniques are combined than when they are used separately.

Hypothesis 1: The noticing of the target form in the input would be greater for the output subjects than for the nonoutput subjects.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Hypothesis 2: The noticing of the target form in the input would be greater for subjects receiving enhanced input than for those receiving unenhanced input.

Hypothesis 3: The noticing of the target form in the input would be greater for +O+IE subjects than for +O–IE subjects and –O+IE subjects.

The fourth through sixth hypotheses are concerned with the acquisition issue. If, as Schmidt (1990, 1995, 2001) claimed, noticing of the form is required for learning that form, it should follow that greater noticing of the target form induced via output, input enhancement, or a combination of both, should lead to increased learning of the form.

Hypothesis 4: Output subjects would demonstrate greater learning of the target form than would nonoutput subjects.

Hypothesis 5: Subjects receiving enhanced input would demonstrate greater learning of the target form than would subjects receiving unenhanced input.

Hypothesis 6: +O+IE subjects would demonstrate greater learning of the target form than +O–IE subjects and –O+IE subjects.

Hypothesis 7 is postulated based on the logical inference that, if the underlying variable for the treatments is the focused attention given to the formal features in the input (i.e., noticing), how it is induced—via output or input enhancement—should make little difference in the learning outcomes as long as noticing is somehow induced.

Hypothesis 7: By virtue of their attention-drawing effects, the two conditions of +O–IE and –O+IE would produce a comparable amount of noticing and learning.

Subjects

Subjects were recruited from among students enrolled in the ESL programs at two major U.S. universities that were similar in both their programs and their enrolled ESL student population. The selection of the subjects was determined on the basis of a test of English relativization administered to intermediate-level classes at the two institutions. This screening test also served as the pretest for the participating subjects. Following Doughty (1988, 1991), the present study chose subjects who demonstrated emerging knowledge of relativization but who had yet to acquire it beyond the initial subject relativization phase. Specifically, students who demonstrated ample knowledge of the target structure or who failed to show any sign of knowledge of grammatical embedding necessary for relativization were eliminated from the subject pool at this point. An arbitrary cutoff score of 80% on the total pretest score was set to exclude subjects from the former category. Students were determined to have had at least rudimentary knowledge of relativization if they (a) produced targetlike subject RCs at least once in the production tests, (b) produced other types of RCs (i.e., direct object or OPREP types) in a targetlike way or in an IL

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

form (mostly with pronoun copies), or (c) responded to the items on the subject RC in the reception tests at better than chance scores.

Out of 110 students who took the pretest, 77 satisfied the learnability requirements. Of these, 47 agreed to participate in the study and completed all the treatments and the posttest. Fourteen randomly selected students who satisfied the above requirements served as the control group and participated only in the pre- and posttests. Thus, including the control group, the number of the final pool of subjects was 61. To ensure pretreatment equivalence of the subjects in their knowledge of English relativization, subjects were assigned to different groups using a stratified random assignment procedure and the pretest results as the basis of stratification. This resulted in the following grouping: 11 for the +O+lE group; 12 each for the +O–lE, –O+lE, and –O–lE groups; and 14 for the control group. To summarize the subject information for all groups combined, the subjects' L1s included Arabic (24), Chinese (6), French (2), Japanese (3), Kazah (1), Korean (11), Persian (1), Polish (1), Portuguese (1), Spanish (6), Thai (4), and Turkish (1). There were 34 males and 27 females. The subjects' mean age was 26.2 years ($SD = 8.22$). Their mean length of residence in the United States was 6.3 months ($SD = 6.17$). Almost all subjects had completed their high school education prior to coming to the United States. The mean length of English instruction was 6.6 years ($SD = 3.6$). All of these factors were divided roughly equally in each group.

Experimental Schedule

The experimental schedule of the study is as follows. First, the researcher visited each class and administered the pretest. The test was subsequently scored for initial screening. The mean interval length between the pretest and the commencement of the treatment for the subjects was 7.13 days. The treatment, conducted in the written mode, consisted of six sessions and took place over a 2-week period. During this period, subjects were requested to come to the computer laboratory outside of class hours and complete the "lesson of the day." The researcher, present at the lab throughout the treatment, supervised and assisted subjects with procedural matters. After signing a consent form and filling out a background questionnaire, subjects engaged in a practice session, the purpose of which was to familiarize them with the activities that were done mainly on the computer and with production work done using paper and pencil. Subjects were not told that this first session was a practice session. The target sessions started with the second session. The topics of all texts were related to relationships and marriage. In the few cases in which subjects could not come to the session on six different days, they were allowed to complete two sessions in one day. After the treatment period, the researcher visited the students' classes once again and administered the posttest. The mean interval length between the last day of the treatment and the posttest administration was 3.43 days. After the posttest was completed, all the treatment subjects filled out a posttreatment questionnaire.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

In an attempt to control for the outside exposure to the target form, the teachers of the participating classes were requested not to teach RCs and not to answer any questions about RCs from their students during the experimental period; all teachers accepted and followed this request. Subjects were asked not to discuss the treatment activities with other people. As a further check, a posttreatment questionnaire probed whether any subjects consulted with anyone or anything about the treatment activities and whom or what they consulted if they did. The analysis of the questionnaire data and the corresponding analyses of the test scores did not reveal any obvious differences in the subjects' performance in this regard (see Izumi, 2000, for details).

Treatment

Although it is assumed that L2 learners are equipped with psycholinguistic mechanisms that underlie the Output Hypothesis (e.g., grammatical encoding, monitoring, noticing mismatches; see Izumi, 2000), the earlier discussion suggested that not all production circumstances provide ideal grounds on which to promote sensitization to language forms. In particular, the capacity limitations of human attentional resources are likely to affect the efficiency of the encoding and monitoring processes in production.[2] Such effects, moreover, are likely to be more pronounced for language learners than for mature L1 users, owing to the greater need of the former group to exercise controlled processing that requires attentional control (Kormos, 1999). Precisely because of the limited and selective nature of attention and the availability of an escape hatch from focus on form in the production process, it is important to ask not only whether output can enhance learning in general terms but also under what conditions it contributes to language learning. The treatment tasks described in the following subsections aimed to create an optimal condition for language learning.

Text Reconstruction Task: Rationale for Use and Major Design Features. The task used in this study was a modified version of a text-reconstruction task used in previous studies (Izumi & Bigelow, 2000; Izumi et al., 1999). Several reasons underlie this decision. First, a reconstruction task by its very nature has the potential for promoting comparisons between the IL output and the TL input. It is essentially a meaning-based pedagogical activity that allows learners to devote some attentional resources to form and that provides both the data and the incentive for the learners to make IL–TL comparisons (Thornbury, 1997). It is, in a sense, a unique "linguistic problem-solving task" (Brett, 1994, p. 332). These are important features of the task in light of the notions of focus on form, in which integration of form and meaning is emphasized. More specifically, one advantage of the reconstruction task lies in its control over the content and form that learners produce. Maximizing the equivalence between the learners' output and the target input should promote direct comparisons between their IL-output and the TL-input forms. Further-

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

more, a reconstruction task that provides a second opportunity for exposure to the original input and for reconstruction permits the investigation of a shift of learners' noticing from the first to second input exposures and their uptake of the form from the first to second outputs.

Overall, the following features characterized the design of the reconstruction task used in this study. First, the entire reading-based treatment was computerized using the Libra program (Farris, 1993), which permitted both a convenient data-collection schedule and a strict control over the time of input exposure and sequence of the activities. Second, as a major modification on the previously used reconstruction task, the input texts were divided into several shorter, semantically coherent subsections to lighten the processing load on the learners. Each subsection was made short enough for subjects to remember its content but long enough to make verbatim memorization difficult. For the most part, each subsection consisted of three sentences. Third, the input texts, which were specially created for this study, contained numerous examples of the target form used in contexts, thereby creating a so-called input flood condition for all groups (see Trahey & White, 1993).[3] Fourth, notetaking was required of subjects as they read the input text. Notetaking was chosen as an online measure of noticing in this study due to its relative advantage in terms of task compatibility (i.e., the measure does not interfere with the reading task) and the preciseness of the measure (i.e., the degree to which the measure excludes items that are not attended to), despite its potential weakness in completeness (i.e., the extent to which the measure includes all items that are attended to). Given that no known measure of noticing (e.g., think-aloud, underlining, or retrospective interview) is perfect in all relevant criteria, it was decided that the nonintrusive, task-compatible measure of notetaking would be appropriate to tap the noticing phenomenon in this study. Reconstruction was also used as another measure of noticing to partially compensate for the weakness of the notetaking.

Treatment Procedures. The goals of the task, as conveyed to the subjects at the beginning of the treatment, were (a) to read and understand the text, (b) to reconstruct the text as accurately as possible (for the output groups) or to answer questions (for the nonoutput groups), and (c) to demonstrate comprehension by writing a recall summary in their L1. Thus, comprehension of the reading text was emphasized for all groups. Subjects were informed in advance that they would have only a limited amount of time to read the given section and could not go back to earlier sections; output-group subjects were also told that they could not go back to their previous productions. Each of the four treatment conditions followed the sequence depicted in Figure 1. Subjects in all groups were first given the entire text and were instructed to read it to get the gist of the text. They were given 1 minute for this skimming phase. During input exposure 1, they were presented with the first subsection of the text for closer reading. At this point, subjects in the two output groups were instructed to "take notes of any and every word" that they thought was "par-

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

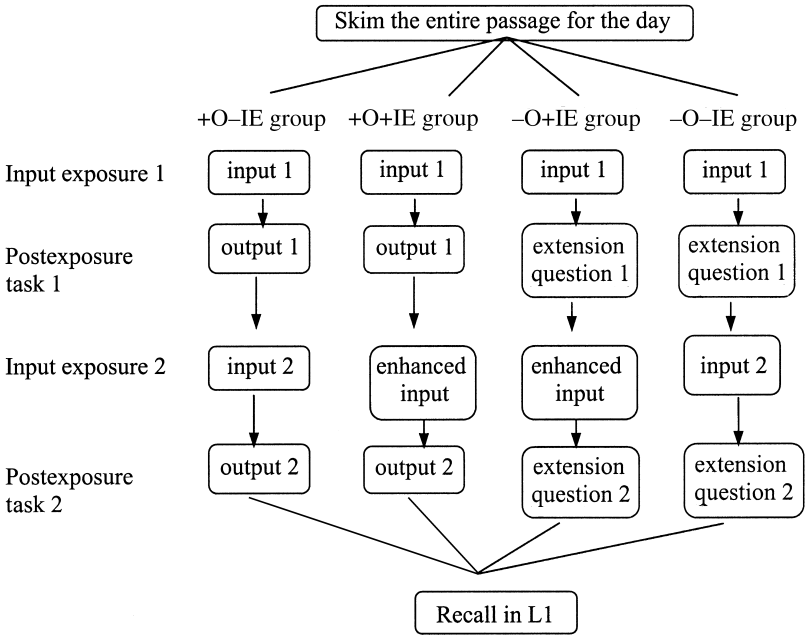


Figure 1. Treatment sequence for each group.

ticularly important, necessary, or useful to reconstruct the text as accurately as possible." Subjects in the two nonoutput groups were instructed to "take notes of any and every word" that they thought was "important, necessary, or useful to understand the text." All of the groups were given note sheets to use for notetaking purposes. To discourage copying of all words in the text, subjects were told not to write down entire sentences but rather to take notes of only those words that they found important, necessary, or useful for their respective purposes. The control for time of exposure was also set in such a way that direct copying was not feasible.

During postexposure task 1, subjects in the output groups were instructed to reconstruct the subsection as accurately as possible in both content and grammar by writing on the reconstruction sheets. They were allowed to use the notes they had taken to help them with their reconstruction. Subjects in the nonoutput conditions, on the other hand, were presented with an extension question that asked for their opinion related to the topic discussed in the text. Extension questions were used in lieu of comprehension questions to prevent inadvertently drawing the learners' attention to form, which was a design problem found in previous studies (Izumi et al., 1999; Izumi & Bigelow, 2000). Furthermore, the multiple-choice format of questions prevented inadvertent production of the target form by nonoutput subjects. An example of a subsection of the reading text and ensuing extension questions is presented

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

in (1). Question 1 followed the first exposure to the text; question 2 followed the second exposure.

(1) He arrived one hour late for the date last Friday. He said that the business meeting which he participated in lasted a long time. Furthermore, the paper which he was working on at the office took him a long time to finish.
Question 1: Do you show up for a date on time?
1. Always 2. Usually 3. Sometimes 4. Never
Question 2: Have you ever made up reasons for being late for a date?
1. Always 2. Often 3. Sometimes 4. Never

No time limit was set for either reconstructing for the output groups or answering extension questions for the nonoutput groups.

After putting their note sheets (and reconstruction sheets for output subjects) in the envelopes provided, subjects were shown the same subsection of the text a second time (input exposure 2) and were directed to take notes on another note sheet for their respective purposes. The enhanced groups were presented with the enhanced text, whereas unenhanced groups were given the same unenhanced text. Subjects in the enhanced condition were directed to attend to the underlined (i.e., enhanced) portion of the text to examine how the complex NPs were formed. To increase the salience of the target form in the enhanced input, the head noun and the relative pronoun were highlighted by a combination of bolding, shadowing, and different fonts and font sizes. The preposition in the RC was also similarly highlighted to indicate the underlying site from which the relative pronoun was extracted. Additionally, the whole NP containing the RC was underlined to make the identification of the complex NP easier.

During postexposure task 2, the output subjects were asked to reconstruct the given subsection on another reconstruction sheet, whereas the nonoutput subjects answered another extension question. As soon as they completed their respective tasks, they were told to put their note sheets (and reconstruction sheets for output subjects) in the given envelopes.

Each step was then repeated for each subsection until the entire text of the day was covered. On completion of the entire text, subjects were directed to write a recall summary (in their L1) on their recall-summary sheets for the day's material. The L1 was used here to further control for the L2 production of the target form. This task was untimed.

In general, it took about 30–60 minutes for the output subjects to complete each session, whereas it took about 30–45 minutes for the nonoutput subjects to complete their lessons each day. The longer time spent by the output subjects was due to the generally longer time required to reconstruct the text, as compared to answering multiple-choice questions. It should be emphasized, however, that the time of exposure to the input text was exactly the same for all groups.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Testing Instruments

To assess the subjects' knowledge of English relativization before and after the treatments, four different written testing measures were used. Two of them—a sentence combination test and a picture-cued sentence-completion test—examined the subjects' productive knowledge; the other two tests—an interpretation test and a grammaticality judgment test—aimed at testing their receptive knowledge. These tests were given in the above order. In all tests, three of the six relativization types represented in the Noun Phrase Accessibility Hierarchy (Keenan & Comrie, 1977) were included in equal number: subject (SU), direct object (DO), and object-of-preposition relatives. Half of these items had RCs embedded in the subject position and half in the object position. To standardize the testing procedure, the entire test session was directed by a recorded guide. The exact time allocated for the test, including the time interval between test items, was determined on the basis of the results and feedback obtained from a series of pilot tests.

Sentence Combination Test. The sentence combination test (SCT) was adapted from Doughty (1988, 1991), who had adapted it from Gass (1982). In this test, subjects were directed to combine two sentences in such a way that the underlined words in the first sentence would be specified or identified by using the information contained in the second sentence (e.g., *The student went downtown* + *I talked to you about the student* → *The student who I talked to you about went downtown*). Subjects were told to begin their combined sentence with the first sentence and not omit any information contained in the two sentences. They were prohibited from using any coordinating conjoiners such as *and*, *because*, *so*, or *when*. The test included 18 items and was completed in 15 minutes.

Picture-Cued Sentence Completion Test. The picture-cued sentence completion test (PCSCT) was also adapted from Doughty (1988, 1991), who had adapted it from Hyltenstam (1984). Unlike these earlier studies, this test was used in the present study as a written elicitation measure for logistical reasons. For each item in this test, subjects saw a picture and heard a sentence that described what was happening in each picture in a simple declarative form (e.g., *The man kisses the woman*). They were then asked to complete a sentence that identified the person in the picture, starting with a prompt (e.g., *Number X is the woman . . .*). To further encourage the use of particular relativization types, subjects were told to follow the instruction about the focus for each sentence (e.g., "focus on what the man does to the woman") and not to use the word *and* in their sentences. The test consisted of 12 items, with 15 seconds to complete each item.

Interpretation Test. The interpretation test (IT) was specially created for this study and was inspired by several sources: the interpretation-type tests used in a series of studies by VanPatten and his colleagues (see VanPatten,

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

1996, for summaries of these studies); studies conducted within the framework of the competition model, which underscores the importance of competition of different cues for sentence interpretation (e.g., Bates & MacWhinney, 1989; MacWhinney, 1987); and L1 research on children's comprehension strategies, which found children's predominant use of the canonical order strategy or the conjoined-clause analysis in interpreting RC sentences (e.g., Roth, 1984; Tavakolian, 1981). In each item of this test, subjects heard a sentence (e.g., *The cat bites the bird which chases the mouse*) that was also written on the test sheet. They were then given 15 seconds to decide which picture best described the sentence among the three choices given. The incorrect alternatives depicted some of the most common interpretations possible for the sentence, which were carefully chosen based on the results of L1 acquisition studies on English RCs previously mentioned (for closer analyses of the data obtained in this test, see Izumi, 2000). The test consisted of 18 items.

Grammaticality Judgment Test. The grammaticality judgment test (GJT) was adapted from Doughty (1988, 1991) and consisted of 36 sentences. Each sentence was presented via audiotape and was also written on the test sheet. Subjects were given 15 seconds to decide whether a given sentence was correct or incorrect. If they decided that the sentence was incorrect, they were asked to correct the error. Following Gass (1982), four error types were included in the incorrect items: pronoun retention, nonadjacency, incorrect relative-marker morphology, and inappropriate relative-marker omission. Subjects were told that the *who–whom* distinction was not important for the purpose of this test.

For the posttest, the same format was followed except that test items were shuffled for all but the PCSCT, and words were substituted for other equally simple words for the GJT and the SCT. Only these minor changes were made because it was felt that more than 2 weeks of interval between the two testing sessions were sufficient to counteract any possible practice effect, which was checked and confirmed by examining the performance of the control group (see the section entitled "Results").

Scoring and Analyses

To measure noticing of the target form, scores were computed by tallying the number of various categories of words written down on the subjects' note sheets and reconstruction sheets, deriving note scores and reconstruction scores, respectively. Note scores were computed for the following categories of words: words inside the relativized NP, words outside of it, content words, function words, head nouns, relative pronouns, and prepositions that accompany the predicate of the clause (e.g., *look for*). The analysis of the reconstruction was carried out by tallying the number of the following items found in the reconstruction: any type of RC attempted, head nouns, relative pronouns, prepositions associated with the predicate of the RC, and the correct use of

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

the target RC. Because of space constraints, and in view of the overall similarity of the results obtained in each of these analyses, this paper reports on only the results pertaining to the three RC words (for notetaking) and to the attempts and targetlike use of RCs (for the reconstruction).

All of the test data were scored by giving one point for a correct response and zero for an incorrect response. The data from each testing measure were scored separately and were then combined to obtain a composite test score. Scoring was also done separately for three different RC types and for different RC positions in the matrix clause, though these results are not reported here (see Izumi, 2000). For the production tests, only the production of the RC that was targeted in the given item was considered correct. Errors related to articles, tense, and spelling were ignored as long as they did not pertain to the formation of the RC. Because omission of the relative pronoun is permissible in the DO and OPREP types of RCs (but not for the SU type), it was considered correct as long as everything else relevant to RC formation was correct. For the IT, answers were scored as either a correct choice or an incorrect choice.

Given the problems a GJT can pose, such as the uncertainty of why a subject judged a sentence to be ungrammatical (see Sorace, 1996), the following procedure was used. First, the subjects' correction of errors was examined to decide whether each item was judged on the basis of the RCs. If the corrections indicated that subjects were judging the test sentences for reasons other than RC-related problems (e.g., tense, articles, and prepositions), these items were determined to be nonapplicable and were eliminated from further analysis. RC-related corrections typically involved movement of the RC to a position adjacent to the head noun if it was not already there, change of word order inside the RC, change of relative pronouns, insertion of a relative pronoun when there was none (applicable only for the subject relativization), and insertion or deletion of pronoun copies. Any items that were not judged or that were judged incorrect but without corrections were excluded from the analysis. The remaining items were then scored for the correctness of the judgment. Owing to the elimination of some test items, which was caused by subjects' failure to complete the test in time or to follow the specified procedure (e.g., not providing corrections in the GJT), test scores were calculated for each subject by dividing the total correct scores by the total number of applicable items. Each subject, therefore, received a percentage score for each test.[4]

RESULTS

Noticing Results

Results of the Note Scores. Table 2 displays descriptive statistics on the number of the three RC words noted by the four treatment groups. This table displays the mean number of each category of words noted during the first and second input exposures as well as the difference (i.e., increase) between

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Table 2. Descriptive statistics for note scores: Head nouns (HN), relative pronouns (RP), and prepositions (Prep) noted per text

		Output								
		Required			Not required			Total		
Input enhancement		1st input	2nd input	Mean diff.	1st input	2nd input	Mean diff.	1st input	2nd input	Mean diff.
Enhanced										
HN	*M*	3.86	5.15	1.29	2.61	4.99	2.38	3.21	5.06	1.86
	SD	1.62	1.74	1.79	1.67	3.30	2.51	1.73	2.61	2.22
RP	*M*	2.66	4.29	1.64	0.40	2.83	2.43	1.48	3.53	2.05
	SD	2.27	2.27	1.95	0.40	3.40	3.31	1.94	2.94	2.73
Prep	*M*	2.78	4.27	1.49	1.74	3.54	1.80	2.24	3.89	1.65
	SD	1.83	2.31	1.54	1.01	2.60	2.07	1.52	2.44	1.77
Nonenhanced										
HN	*M*	2.74	3.02	0.28	3.11	3.23	0.12	2.93	3.12	0.20
	SD	1.11	1.40	1.27	1.63	2.11	0.99	1.38	1.75	1.12
RP	*M*	0.54	1.19	0.65	0.65	0.78	0.13	0.60	0.99	0.39
	SD	0.82	1.17	0.66	0.92	1.12	0.49	0.85	1.14	0.63
Prep	*M*	2.15	2.38	0.22	2.00	2.23	0.23	2.08	2.30	0.23
	SD	1.40	1.63	0.77	1.48	1.85	0.98	1.41	1.71	0.86
Total										
HN	*M*	3.27	4.04	0.76	2.86	4.11	1.25			
	SD	1.46	1.88	1.59	1.64	2.86	2.19			
RP	*M*	1.55	2.67	1.12	0.53	1.81	1.28			
	SD	1.96	2.35	1.48	0.70	2.69	2.61			
Prep	*M*	2.45	3.28	0.83	1.87	2.89	1.02			
	SD	1.62	2.16	1.34	1.24	2.31	1.74			

Note. Mean difference values for *M* and *SD* were obtained by subtracting the value of the first input from that of the second input for each individual subject and then dividing the total of all subjects by the number of subjects.

the two. Also displayed in this table are the results for the output versus non-output groups (shown at the bottom) and for the input enhancement versus nonenhancement groups (shown at the far right) to facilitate the comparison along these two independent variables of the study.

An examination of the noting patterns of the three RC words revealed that the two enhancement groups had a greater increase from the first to second exposures in the number of each of these words noted than did the nonenhancement groups. These data were submitted to a $2 \times 2 \times 2 \times 3$ ANOVA, with output and input enhancement entered as the two between-subjects factors and time and word type entered as the within-subjects factors. The ANOVA revealed significant main effects for input enhancement, $F(1, 43) = 10.84$, $p < .01$, time, $F(1, 43) = 18.36$, $p < .001$, and word type, $F(2, 86) = 124.81$, $p < .001$, as well as significant interaction effects between time and input enhancement, $F(1, 43) = 9.64$, $p < .01$, word type and input enhancement, $F(2, 86) = 8.40$, $p <$

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

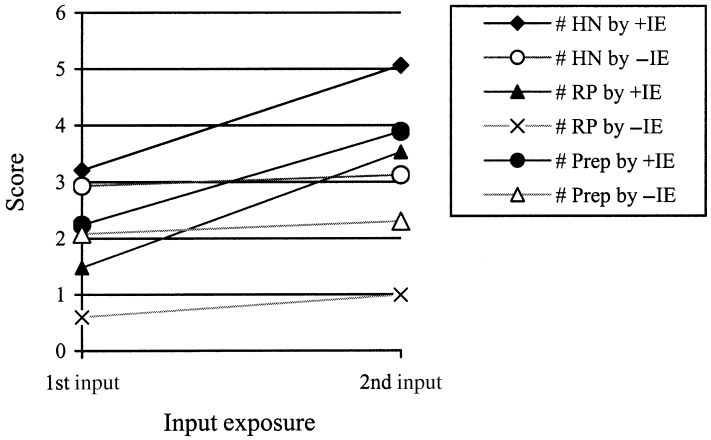


Figure 2. Mean note scores for the highlighted RC words by the ± input enhancement groups.

.001, time and word type, $F(2, 86) = 13.47$, $p < .001$, and time, word type, and input enhancement, $F(2, 86) = 8.45$, $p < .001$.

To identify the sources of the three-way interaction involving time, word type, and input enhancement, an interaction plot was examined for the number of three RC words noted by the enhancement and nonenhancement groups at the two input exposures (see Figure 2). A visual inspection of the figure revealed that the enhancement groups experienced an increase from the first to second exposures in all three types of words, whereas the nonenhancement groups did not change much for any of these words. Additionally, a post hoc analysis of the word-type effect using paired *t*-tests (with an adjusted alpha level of $p < .17$ for three-way comparisons) revealed that the three RC words were noted in significantly different amounts. Head nouns were noted significantly more than prepositions: 3.57 versus 2.62, $t(46) = 13.84$, $p < .001$. Prepositions were noted significantly more than relative pronouns: 2.62 versus 1.63, $t(46) = 5.64$, $p < .001$. Though not strikingly different, a relatively greater increase in the number of relative pronouns noted than for the other two words (see Table 2 also) may have contributed to the significant three-way interaction.

Within-group comparisons using paired *t*-tests generally confirmed these observations. A significant increase in the number of head nouns was not observed for the nonenhancement groups although it was observed for the two enhancement groups: for +O+IE, $t(10) = 2.39$, $p < .05$; and for –O+IE, $t(11) = 3.28$, $p < .01$. For relative pronouns, a significant increase was found for +O+IE, $t(10) = 2.79$, $p < .05$, for +O–IE, $t(11) = 3.41$, $p < .01$, and for –O+IE, $t(11) = 2.53$, $p < .05$. For prepositions, a significant increase was observed only for the two

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Table 3. Mean reconstruction scores for the number of attempts at RCs and targetlike use of OPREP relatives by the two output groups

	First output		Second output		Mean diff.	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
+O+IE group						
Attempts	5.29	1.26	8.04	0.60	2.75	1.10
Targetlike use	2.07	1.46	4.38	1.74	2.31	0.97
+O–IE group						
Attempts	4.97	1.92	6.73	2.08	1.77	1.13
Targetlike use	2.02	1.78	3.58	2.34	1.57	1.25

Note. Mean difference values for *M* and *SD* were obtained by subtracting the value of the first output from that of the second output for each individual subject and then dividing the total of all subjects by the number of subjects.

enhancement groups: for +O+IE, $t(10) = 3.21$, $p < .01$; for –O+IE, $t(11) = 3.11$, $p < .05$.

Thus, it seems that the groups that received visually enhanced texts took a greater number of notes of the RC words as they engaged in the reading or reconstruction tasks. To the extent that notetaking served as an adequate measure of noticing, these results indicate a positive role played by the visual input enhancement. In comparison, the effect of output was not evident here. The next section focuses on the reconstruction produced by the two output groups during the treatment.

Results of the Reconstruction Scores. Table 3 displays the mean reconstruction scores for the number of attempts at the RC and the targetlike use of the OPREP relatives. Figure 3 shows these scores graphically.

Both groups increased in the number of attempts at RCs and the targetlike use of the OPREP relatives from the first to the second output. The trend seems to be for the +O+IE group to improve more than for the +O–IE group in both the attempt and the targetlike use. In terms of the percentage of the number of attempts out of the total number of RCs included in the input texts, 56% and 80% of the RCs were found to have been attempted by subjects in both groups in the first and second reconstructions, respectively. For targetlike use, 22% and 43% of the RCs embedded in the texts were reconstructed in a targetlike manner in the first and second reconstructions, respectively. If the amount of targetlike use is considered out of the number of attempts at the RC, there was an average of 40% successful use of the target relatives in the first reconstruction and 54% in the second reconstruction.

The data for the number of attempts and targetlike use were submitted separately to two-way, mixed design ANOVAs using group (+O+IE vs. +O–IE) as a between-subjects factor and time (first vs. second output) as a within-subjects factor. For the number of attempts, the ANOVA revealed a significant main effect for time, $F(1, 21) = 94.40$, $p < .001$, and a significant interaction between

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

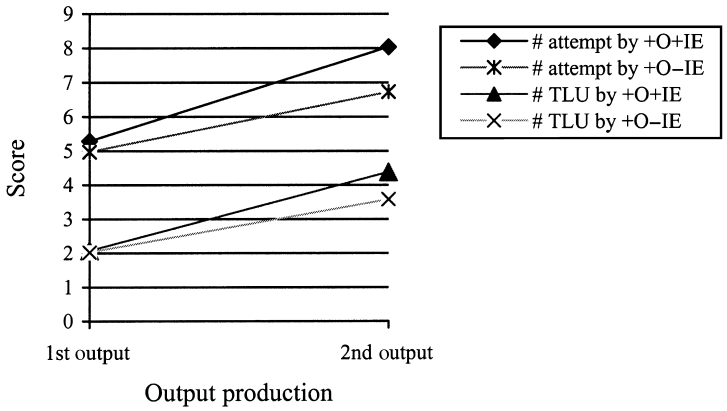


Figure 3. Mean reconstruction scores for the number of attempts at RCs and targetlike use of the OPREP relatives by the two output groups.

time and group, $F(1, 21) = 4.44$, $p < .05$. Figure 3 reveals that the interaction effect was caused by the greater increase over the two reconstruction opportunities for the +O+IE group than for the +O–IE group. For the targetlike use, the ANOVA revealed a significant effect only for time, $F(1, 21) = 67.92$, $p < .001$, which suggests that subjects improved in their accurate use of the target relatives regardless of the enhancement factor.

To summarize, although the enhancement seemed to have some positive impact on the number of attempts at the RCs, in general the two output groups behaved similarly in their overall success at reconstruction, which may be taken to suggest that the two output groups did attend to the target form in the input and in their output, attempting and succeeding in producing a fair number of RCs in their reconstructions.

Test Results

Turning now to the test results, Table 4 displays the descriptive statistics for the total test scores on the pretest, posttest, and the gains made from the pre- to posttests by each treatment group. Shown in the bottom rows are the results pertaining to output versus nonoutput groups. The columns on the far right show the results for the enhancement versus nonenhancement groups.

Before examining the effects of the treatments, pretreatment equivalence of groups in their knowledge of English relativization was checked by submitting the pretest scores to a two-way multivariate analysis of variance (MANOVA) in which output and input enhancement were used as the two between-subjects independent variables. The data were also submitted to a one-way MANOVA comparing the pretest scores across the five groups (including

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Table 4. Descriptive statistics for total test scores

Input enhancement	Output								
	Required			Not required			Total		
	Pretest	Posttest	Gain	Pretest	Posttest	Gain	Pretest	Posttest	Gain
Enhanced									
M	54.42	68.17	13.75	50.84	56.24	5.40	52.55	61.94	9.39
SD	12.25	14.71	9.83	16.48	18.71	7.23	14.40	17.62	9.41
Nonenhanced									
M	50.99	64.91	13.93	54.06	59.44	5.38	52.53	62.18	9.65
SD	14.41	16.11	9.41	14.32	15.86	5.74	14.14	15.88	8.79
Total									
M	52.63	66.47	13.84	52.45	57.84	5.39			
SD	13.23	15.19	9.38	15.19	17.04	6.41			

Note. Gain values for *M* and *SD* were calculated as in Tables 2 and 3. The results for the control group are: pretest ($M = 53.82$; $SD = 15.17$); posttest ($M = 56.99$; $SD = 13.82$); gain ($M = 3.17$; $SD = 6.15$).

the control group). The results of these MANOVAs indicated no significant difference among the groups. Therefore, any difference that may be observed on the posttest or on the pre- to posttests gains is attributable to the respective treatments.

The overall test results displayed in Table 4 reveal that the two output groups experienced greater improvement on the posttest than did the two nonoutput groups. To test for the statistical significance of these differences, the gain scores were submitted to a two-way MANOVA. The results revealed a significant main effect for output: Wilks's Lambda .696 and Pillai trace .304, $F(4, 40) = 4.36$, $p < .01$. However, there was no significant main effect for input enhancement and no significant interaction effect. This indicates that subjects engaged in output–input activities gained significantly more on the tests of relativization ability when compared to subjects engaged in input comprehension activities. The results also indicate, though, that subjects exposed to the enhanced texts did not demonstrate significantly better gains on the test as compared to those exposed to the unenhanced texts. A one-way MANOVA comparing the gains of the five groups (including the control group) also revealed similar statistical results in that a significant main effect was found for group: Wilks's Lambda .569 and Pillai trace .482, $F(16, 162) = 2.06$, $p < .05$. A Fisher's PLSD test indicated that the +O+IE and +O–IE groups outperformed the –O+IE, –O–IE, and control groups, whereas there was no significant difference between the two output groups, between the two nonoutput groups, or between the two nonoutput groups and the control group. Thus, it seems that the output–input treatment had a significant impact on the learning of English relativization, whereas input enhancement and a comprehension-based treatment failed to induce an equally beneficial effect on learning.

As a further test for the changes made by different groups, paired *t*-tests

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

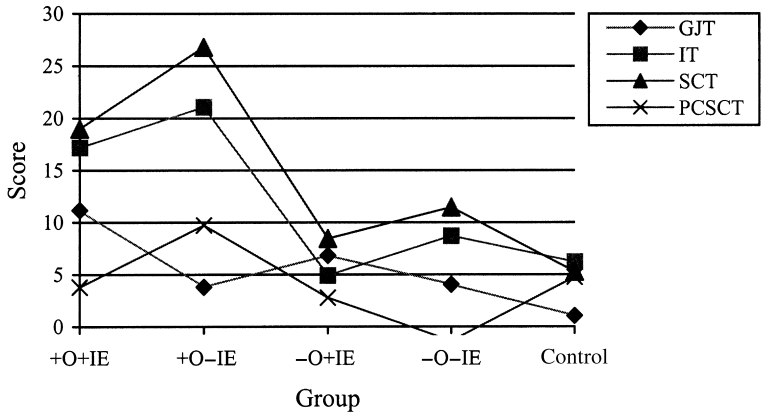


Figure 4. Mean gain scores on the four testing measures by the five groups.

were performed on the total pre- and posttest scores for each group. These tests revealed significant improvement on the relativization ability over the two testing sessions for all groups except for the control group: for +O+IE, $t(10) = 4.64$, $p < .001$; for +O–IE, $t(11) = 5.13$, $p < .001$; for –O+IE, $t(11) = 2.57$, $p < .05$; for –O–IE, $t(11) = 3.24$, $p < .01$; and for the control group, $t(13) = 1.93$, $p = .076$. The fact that not only the two output groups but also the two nonoutput groups improved significantly seems to indicate some positive impact of the exposure to the input flood during the treatment.

To examine the effectiveness of the treatment, effect sizes were computed. The effect size estimates the magnitude of change attributable to instructional treatments and gives us valuable information as to the practical importance of the treatment effect without regard to the sample size of the study (Kramer & Rosenthal, 1999). Following a recent study of meta-analysis of L2 instruction studies by Norris and Ortega (2000), Cohen's *d* was used to calculate the effect sizes for the experimental groups as compared to the control group. Using the composite gain score means, the following effect sizes were obtained: 1.12 for +O+IE, 1.14 for +O–IE, 0.34 for –O+IE, and 0.37 for –O–IE. When these values are compared with the mean effect size observed in Norris and Ortega's study for treatments of the kind that the current study adopted (i.e., focus-on-form implicit treatments: $d = 0.69$), we can see that the figures obtained for the present output–input treatment are quite large, whereas input-only treatments had a small magnitude of effect.[5]

The MANOVA results previously reported indicated a significant main effect for output. This prompted further statistical analyses on this variable. Consequently, univariate analyses of variance (ANOVAs) were performed on the gain scores obtained from each testing measure to determine in which test(s) the significant effect for output was obtained (see Figure 4). The ANOVAs revealed

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Table 5. Summary of the results

Hypothesis	Prediction	Results
H1	Noticing of the target form: +output > –output	Not supported
H2	Noticing of the target form: +enhancement > –enhancement	Supported
H3	Noticing of the target form: +O+IE > +O–IE, –O+IE	Mixed; on note score, +O+IE, –O+IE > +O–IE; on reconstruction score, +O+IE > +O–IE in RC attempt
H4	Learning of the target form: +output > –output	Supported
H5	Learning of the target form: +enhancement > –enhancement	Not supported
H6	Learning of the target form: +O+IE > +O–IE, –O+IE	Mixed; +O+IE = +O–IE > –O+IE on test gains
H7	Noticing and learning: +O–IE = –O+IE	Not supported; on note score, +O–IE < –O+IE; on test gains, +O–IE > –O+IE

a significant effect for output on the SCT, $F(1, 43) = 7.32$, $p < .05$, and on the IT, $F(1, 43) = 4.99$, $p < .05$. To include the control group, another set of ANOVAs was also run on the same data set using Group as the independent variable. It indicated a significant effect for group for the SCT, $F(4, 56) = 3.48$, $p < .05$. A Fisher's PLSD test revealed that the +O+IE group outperformed the control group, and the +O–IE group outperformed all other groups except for the +O+IE group.[6]

Thus, although not all testing measures showed equally strong discrimination among groups and over the two testing sessions, it appears that the benefits of output were not limited to one mode of relativization ability but manifested in both production and comprehension measures.[7] The summary of the results in terms of the seven hypotheses is presented in Table 5.

Further Analysis: Comprehension of the Reading Texts

One of the most important features of focus on form as advocated by Long (1991) and others (Doughty & Williams, 1998; Long & Robinson, 1998) is its aim at achieving integration of attention to form and meaning as learners engage in meaning-oriented activities. In view of the positive learning outcomes for the output groups, the question may be raised as to whether any deterioration of comprehension occurred for the output subjects as a result of their greater attention to and processing of the form. To investigate this question, recall summaries written by all treatment subjects were analyzed using the procedure adopted by Doughty (1988, 1991). In this procedure, the researcher first prepared a list of propositions representing the major ideas contained in each text. Then, native or near-native speakers of the given language examined whether each proposition was understood fully, partially, or not at all.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Subsequently, these evaluations were converted into comprehension scores by giving one point if the proposition was evaluated to have been understood fully, one half point when understood partially, and no point when not understood at all. A mean score was then calculated over all the texts that were read and summarized for each subject, yielding an overall comprehension score.[8]

The mean comprehension score obtained for the two output groups was 65.0 (SD = 12.2; for +O+IE, M = 64.4, SD = 10.6; for +O–IE, M = 65.5, SD = 13.9), and the mean score for the two nonoutput groups was 67.5 (SD = 17.0; for –O+IE, M = 60.7, SD = 16.8; for –O–IE, M = 75.0, SD = 14.6). Although there are some between-group differences in the descriptive statistics, a two-way ANOVA revealed no significant main effect for either output or input enhancement and no significant interaction effect; these results were also corroborated by a one-way ANOVA using Group as the independent variable. This indicates that the output subjects obtained their significant gains in learning while maintaining a comparable level of comprehension to that of the nonoutput subjects. The mean comprehension score of 65% for the output groups also suggests that these subjects, like subjects in the nonoutput groups, achieved a fairly high level of comprehension of the reading passages (especially considering the demanding nature of the recall summaries that test not only the learners' comprehension but also their memory of the passage content). Thus, it may be concluded that the output subjects successfully learned the target form without sacrificing comprehension. In focus-on-form terms, it appears that the output–input treatment in this study succeeded in achieving an adequate level of integration of attention to form and meaning. Nonoutput groups, on the other hand, achieved a comparable level of comprehension to that of the output subjects yet failed to demonstrate as much learning of the form. This seems to corroborate findings of earlier studies that comprehensible input or even comprehended input is not sufficient to ensure automatic learning of the form embedded in the input.

DISCUSSION

Three major findings are evident in this study. First, those engaged in the output–input treatment outperformed those exposed to the same input for the sole purpose of comprehension in learning English relativization. Although note-score analyses did not reveal any significant impact of output, the immediate uptake of the form evidenced in the reconstruction suggested that output subjects attended to the target form in the input and in their own output. It was also observed that a relatively high level of comprehension of the input texts accompanied the superior learning by these subjects. Second, in contrast to the positive effect of output, visual input enhancement failed to show any measurable effect on learning. This was so, despite a clear indication in the note scores that the enhancement had a significant impact on the noticing of the target-form items in the input. Third, in view of the above, it did not

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

appear that the effects of output on noticing and learning were comparable to those of input enhancement. In what follows, an integrated account of these findings will be sought in reference to the theoretical underpinnings of—and previous research on—noticing, output, and input enhancement.

The Effect of Output on Learning English Relativization: The Noticing Function of Output

The positive effect of output demonstrated in this study is consistent with the hypothesized function of output in SLA. In particular, the noticing function underscores the interconnectedness of input and output processes in SLA. Based on previous research that indicates insufficiency of comprehensible input in driving L2 development and on research that points to the importance of attention in learning, it was hypothesized that learners' attention somehow needs to be drawn to the crucial form features to promote their learning. Output was considered to be one means to achieve this aim by prompting the learners to find problems in their IL through their production attempt. It was further assumed that, on exposure to relevant input immediately after their production experience, the heightened sense of problematicity would lead them to pay closer attention to what was identified to be a problematic area in their IL. In short, pushed output can induce the learners to process the input effectively for their greater IL development.

The results of the learners' reconstruction lend some support for the attention-inducing effect of output. The test results further indicated superior learning of the form by the output subjects as compared to subjects who were exposed to the same input for the sole purpose of comprehension. That such superior performance of the output subjects was found not only in a production test but also in a comprehension test reinforces the contention that the output treatment affected the subjects' IL knowledge, not just their use of some production strategies. Both the process and product measures used in this study, therefore, lend some empirical evidence for the noticing function of output with respect to the acquisition of English relativization by adult L2 learners.

Reexamination of Noticing and Its Relationship to Input Enhancement, Output, and SLA

In contrast to the positive results of output, why did visual input enhancement fail to induce greater learning of the form? It was previously argued that one of the differences between input enhancement and output is that, whereas attention in the former is induced by external means, attention in the latter arises internally through production processes. Interpreting the present results in this light, it appears that these two approaches to focus on form do not promote learning with an equal level of efficacy even though they both supposedly involve learner attention as the crucial variable. Out-

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

put, as an internal priming device, clearly worked better to promote learning in this study.

Another difference between output and visual input enhancement may be that pushed output, by virtue of producing utterances, can place the learner in an ideal position to make a cognitive comparison between the IL and TL forms (Doughty, 2001; Nelson, 1987; Saxton, 1997a, 1997b), whereas this may not be the case with input enhancement. As Saxton's Contrast Theory claims, SLA, as with L1 acquisition, may be promoted when the learners' error in their output is followed immediately by a juxtaposed targetlike form that shows contrast to the learners' preceding utterance. This, Saxton claimed, may create a favorable condition for the learners to make a cognitive comparison between their IL form and the TL form, leading them to expunge the nontargetlike form from their developing IL in favor of the TL form. Recent studies on the role of recasts in SLA show that L2 learners can indeed benefit from being exposed to the targetlike use of form in discoursally contingent positions (e.g., Doughty & Varela, 1998; Long, Inagaki, & Ortega, 1998; Mackey & Philp, 1998).

It is possible that the cognitive comparison may be induced by visual input enhancement as well. However, the enhancement, which is solely concerned with drawing the learners' attention to form, does not necessarily encourage further cognitive processing that may be necessary for acquisition. This condition of input enhancement might render form learning essentially a hit-or-miss affair, with only some of the learners likely to benefit fully from it (perhaps those with most form-conscious tendencies, those with metalinguistically sophisticated prior knowledge, or both). Thus, the contrasting results of the output and the input enhancement in this study may be explained by positing that the input enhancement was not sufficient to induce noticing of the IL–TL mismatches. Output, in contrast, promoted both the processes of noticing the form and noticing the mismatches, which enabled the output learners to attain successful and superior learning of the form. With these general accounts as a backdrop, further illumination of the learning processes is sought in the following section through explanations of a more psycholinguistic nature.

Amount of Processing versus Levels, Depth, and Type of Processing

The lack of a significant advantage of enhancement on learning was particularly striking on consideration of the fact that there was strong evidence in the notetaking measure that the enhancement subjects paid attention to various form elements in the input. To find a psycholinguistically informed account for these results, the basic logic behind the hypothesized effect of input enhancement needs to be reexamined. The logic is generally set up in two steps. First, the perceptual salience created by highlighting the input will draw the learner's attention to the highlighted forms. Second, once the first step is

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

successful, learning of the attended form will occur based on the premise that attention is what mediates input and intake.

Regarding whether the first step in this logic was met in this study, one may argue that subjects in the enhanced groups did not in fact notice the target forms—that is, because of measurement problems associated with notetaking, one cannot assume that taking notes of form-related words equals noticing of these forms. However, the problem with notetaking in this study seems to be with underestimation of actual noticing rather than with overestimation. For example, one can argue that the absence of RC words written as notes does not necessarily mean that no attention was paid to these words, given the incompleteness of the measure. However, it is difficult to argue that the presence of words written down as notes does not mean that at least some attention was paid to these words; it is difficult to imagine a situation where one bothers to write down words without paying any attention to them. This is all the more true when notetaking of these words occurs rather consistently and significantly, as was the case for the enhancement subjects in this study. Given this understanding, then, the significantly greater amount of form-related words noted by the enhancement subjects can be taken to indicate that their attention was indeed drawn to the form. (See also Izumi, 2000, on the responses to the posttreatment questionnaire, which lend further support to this conclusion.) If the minimum requirement of noticing, as defined by Schmidt (1990, 1995, 2001), is to pay attention to key grammatical elements in the input with greater than a threshold level of subjective awareness (i.e., reportable subsequent to the experience), then these data indicate that the basic requirement of noticing has been met, in Schmidt's sense of the term.

The second step in the logic states that success in drawing the learners' attention would lead to learning the attended form. An implicit assumption seems to be that what matters with learners' attention is its focus and quantity; the basic premise is that the more attention to the form, the better its chance of acquisition. The method of notetaking that was used in this study was based on this basic premise (i.e., the quantitative analyses of the notes would reflect the degree of learners' noticing and, thus, learning of the form). As a result of this emphasis on quantitative aspects of attention in the commonly used attentional research paradigm in SLA, what has been left aside is the consideration of the qualitative aspects of attention—namely, what levels and type of attention and processing are engaged with the forms in focus. On the basis of the results obtained in this study, however, it appears that these are precisely the issues that need to be examined more critically.

Levels and Depth of Processing. In cognitive-psychology literature, perception is considered to involve the rapid analysis of stimuli at a number of levels. Preliminary stages are concerned with superficial features, such as physical aspects of the stimuli, whereas later stages are more concerned with matching the input against the already stored knowledge base. In their seminal paper on memory research, Craik and Lockhart (1972) referred to this no-

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

tion of a series of processing stages as depth of processing, where depth is defined in terms of the relative degree of semantic and cognitive analysis and elaboration done on the input stimuli. Craik and Lockhart argued that "analysis proceeds through a series of sensory stages to levels associated with matching or pattern recognition and finally to semantic-associative stages of stimulus enrichment" (p. 675). In this levels-of-processing framework of human memory, the persistence of memory traces is understood to be a function of the depth of analysis, with deeper levels of analysis leading to more elaborate, longer lasting, and stronger traces. Maintaining information at one level of processing by rehearsing it repeatedly or by sustaining continued attention to certain aspects of the stimulus will not, by itself, lead to improved retention unless a shift to deeper levels of analysis occurs.

Considering the findings of the present study within this framework, it is possible that attention evidenced by the enhancement took place at a relatively shallow level without necessarily shifting to deeper and more elaborate processing levels. If so, it should be no surprise, based on Craik and Lockhart's (1972) argument, that the evidence of the large amount of attention to certain form elements alone is not directly related to learning of the new associative connections existing in the input stimuli. In other words, it does not matter how much attention is devoted to the form unless the quality of attention somehow changes to involve deeper and more elaborate processing. What causes deeper levels of processing? Craik and Lockhart argued in this regard that

> the depth at which primary memory operates will depend both upon the usefulness to the subject of continuing to process at that level and also upon the amenability of the material to deeper processing. . . . There are at least three sources of the failure of processing to reach this [deep] level: *the nature of the material, limited available processing capacity*, and *task demands*. . . . Manipulations that influence processing at a structural level should have transitory, but no long-term effects. . . . Long-term recall should be facilitated by manipulations which induce deeper or more elaborate processing. (pp. 679–680; emphasis added)

In the current study, the "nature of the material" refers to the input materials and the target form, both of which were equivalent for all groups of subjects. The target form, moreover, was carefully chosen in light of the learnability considerations so that it would be learnable by the selected subjects. The "limited available processing capacity" was also carefully controlled by presenting only a manageable batch of input sentences at a time, which was also held equivalent for all groups. On the other hand, task demands were varied in the present study along the two independent variables, output and input enhancement. Thus, in this framework of analysis, it could be argued that the input enhancement may have caused mere recirculation or rehearsal at the same, relatively shallow processing level, which led the learners to experience only a short-term retention of the attended form. On the other hand, the

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

greater learning evidenced by the output subjects suggests that output triggered deeper and more elaborate processing of the form, which led them to establish a more durable memory trace.[9]

Type of Processing. Another concept that is closely related to the levels of processing is the notion of type of processing. In the previous SLA literature, type of processing (or of attention) has been discussed in terms of, for example, focal versus peripheral attention, implicit versus explicit learning, intentional versus incidental learning, and noticing versus understanding (e.g., Schachter, 1998; Schmidt, 1990, 1995, 2001; Tomlin & Villa, 1994). Other types of processing important to SLA are discussed by Robinson (1995), who, in defining noticing as detection plus rehearsal in short-term memory, argued that learning outcomes are determined by the level of activation of information in short-term memory, which is caused by rehearsal and elaboration.

As compared to these various conceptualizations of the types of cognitive processing, little attention has been paid in the SLA literature to the notion of integrative processing. Graf (1994) explained it as follows:

> Integration focuses on connections among the units that define an individual item, such as a word, an object, or a sentence; these kinds of connections are formed or strengthened when the subject either perceives coherence among separate stimulus components (e.g., under the guidance of preexisting representations or gestalt laws like proximity or common fate) or conceives a structure for processing target features as a single entity. (p. 685)

Graf and Schacter (1989) argued that so-called unitization, which is a product of integrative processing, does not happen automatically when an item is presented for study but instead depends on study activities that involve conceiving the presented stimuli as coherent entities. The kind of activities that promote integrative processing and unitization were discussed by Graf and Schacter in the context of their investigation—that is, learning or remembering paired words in one's L1 by requiring subjects to generate a sentence that connects the two words in a meaningful manner (e.g., from *book–forest*, one produces *The book included a picture of a forest*). It is argued that semantic, elaborative processing thus invoked "provides the 'glue' (the structure) that unitizes sensory-perceptual and other features of paired words that are processed concurrently during the study trial" (p. 939).

The notion of integrative processing is of great relevance to SLA because it underscores the importance of not only attending to individual form elements that make up a structure but also of perceiving or conceiving the relationship among them to cause structure (grammar) learning. In a study on the roles of memory and attention in SLA, Williams (1999) invoked the notion of integrative processing to account for the results of his study. He argued that knowledge of distributional rules, such as morphological agreement among the determiner, noun, and adjectives in Italian, does not simply emerge out of

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

memory encodings of the relevant forms but rather depends on how attentional resources are appropriately allocated over the relationships among related form elements. In other words, unless one perceives the relationship among related form elements, morphological concordances may never be acquired. In a similar vein, applying the notion of integrative processing in the present study, it may be argued that, to learn English RCs, one needs to not only pay attention to key form elements in the input, such as a head noun, a relative pronoun, and a preposition, but also to focus one's attention on how they are related to one another (e.g., the relative pronoun is associated with the head noun as being coreferential with it; its extraction site is identified). Attending to the individual items like *which* or *who*, no matter how intensely one does so, will not by itself lead to the acquisition of the RC structure unless these items are grasped in relation to other related items in the same clause or sentence.

Furthermore, if semantic or elaborative processing induced via a sentence-generating task prompts unitization in remembering paired words by adult native speakers, it may be argued in the context of the present study that attention to meaning and the semantic processing that was induced in the comprehension task served only as a rough guide directing L2 learners to how the individual form elements are related to one another. On the other hand, output processing, required in the output treatment, pushed these learners further in their cognitive processing and prompted them to perceive or conceive the unitized structure. This occurs by virtue of the grammatical encoding operations performed during production. As a consequence, the output task served effectively both as the stimulator of integrative processing and as the glue to connect individual form elements, which, one might say, were only vaguely related to one another during the comprehension process.

Rethinking the Roles of Output and Input Enhancement in SLA

Based on the foregoing discussion, I posit that the cognitive processing necessary to acquire a complex structure like the English RC involves (a) sensory detection of key form elements such as the head noun and relative pronoun, and (b) further cognitive processing—notably, integrative processing—that connects and organizes related form elements and captures them as a coherent structural set. Following Robinson (1995), noticing can be defined here as sensory detection plus further cognitive processing on the detected data, and these two processing operations together are hypothesized to lead to Schmidt's (1990, 1995, 2001) sense of noticing that is both necessary and sufficient for learning. Cursory and superficial processing of input does not lead to learning of the target structure, no matter how consciously or intensely one attends to a particular form.

To summarize, then, output, especially pushed output, promotes not only detection of forms but also integrative processing to conceive a coherent structure among the detected elements. This, together with the juxtaposition

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

of the IL–TL forms, creates a favorable condition for the learner to notice the mismatches between these two versions. The significant learning of the structure for the output subjects in this study can thus be accounted for. On the other hand, visual input enhancement encourages sensory detection by means of highlighting, but this does not necessarily lead to integrative processing. The automatic link from sensory detection to further cognitive processing cannot be assumed. Seen in this way, the discrepancy between the notetaking results and learning outcomes found in this study may be explained by assuming that the notetaking reflected the learners' detection of individual form elements in the input but not whether or what kind of connections or integrations they have achieved among them. In other words, notetaking of the kind used in this study may have been too gross a measure to capture learners' noticing, if noticing is to mean detection plus further cognitive processing.

This line of argument also finds support in the findings of previous SLA studies on visual input enhancement, which seems to work best in combination with other forms of assistance (e.g., semantic elaboration, as in Doughty, 1988, 1991; a focused production task, as in Williams, 1999; or activation of prior knowledge, as in Shook, 1994) but not so well when used in isolation. It appears that highlighting alone is often not enough to prompt the learners to go beyond simple detection of forms, and additional instructional assistance of some sort may be needed to trigger further cognitive processing, at least for some complex grammatical forms.

To recapitulate, the advantage of pushed output may lie in promoting three related processes that are important to learning complex grammatical forms such as English RCs: (a) detection of formal elements in the input via priming induced by internal feedback and monitoring mechanisms engaged during production processing; (b) integrative processing of the target structure, prompted by the grammatical encoding operations engaged during production processing; and (c) noticing of the mismatches between one's IL form and the TL model, which is fostered by the engagement of these processes and by the juxtaposition of the two versions that highlight any differences between the two uses of the form.

Suggestions for Further Research

Some of the issues remaining to be studied include the following. First, long-term effects of the output–input treatment need to be examined. Second, for both output and input enhancement, possible interactions between the treatment type and form type need further investigation. For this, in addition to consideration of learnability and other factors, such as formal complexity and functional importance attributed to a given form, it will be necessary to consider the kind of cognitive processing that may be needed for the learner to acquire the form. For example, does the form to be acquired need to be processed at the word level (e.g., content words), phrasal level (e.g., adjective–

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

noun agreement in Spanish), clausal or sentential level (e.g., RCs), or suprasentential level (e.g., articles, choice of two verb tenses in a given situation)? The form type may also be considered in terms of whether the learners already possess at least partial knowledge of the form or whether the form is completely new to them. Further investigation into the effectiveness of different combinations of instructional techniques will also be useful. In particular, future studies need to consider carefully what type of enhancement procedures stimulates the right kind of cognitive integration that is necessary for the form to be acquired. Modality is yet another issue: Will the same results be obtained if the modality is changed from the written to the oral or aural modes? The prevalence of oral language use in naturalistic and classroom contexts suggests the importance of tackling this issue.

CONCLUSION

This study set out to investigate whether output and input enhancement, in isolation or in combination, can promote noticing and learning of English relativization by adult ESL learners. The results found in favor of output suggested a three-fold facilitative impact of pushed output on L2 acquisition. That is, under some circumstances, output promotes (a) detection of formal elements in the input, (b) integrative processing of the target structure, and (c) noticing of the mismatches between one's IL form and the TL input. The lack of any significant impact of visual input enhancement also suggested that these advantages of output may not be shared by the superficial external manipulation of the target form in the input, which, without any additional instructional assistance, may help only with detection of the highlighted form items but does not necessarily engage the learner in further cognitive processing. Without denying the essential role of input in SLA, the present study provided empirical evidence for a psycholinguistically motivated role of output in L2 development. It is expected that further research on the roles of output and various forms of input enhancement will be pursued with rigor and increasingly refined methodology so that SLA research can be usefully and effectively applied to language teaching.

(*Received 26 September 2001*)

NOTES

1. Thus, the reconstruction task was used both as a treatment task for the output groups and as a noticing measure. Another measure—a posttreatment, retrospective questionnaire—was also used to probe the attentional focus of the subjects during the treatment. Its results, however, are not reported in this article (see Izumi, 2000, chap. 6).

2. It should be noted that recent theories of attention (e.g., Wickens, 1989) assume that there are multiple pools of resources available in human attentional mechanisms. According to these theories, effective time-sharing or dual-task performance is considered possible if the task draws on different pools of resources (e.g., verbal vs. spatial, as in talking while driving). See Robinson (1995) for a discussion of this issue as applied to SLA.

3. Given that an input flood is a type of input enhancement, one could say that all treatment

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

groups received some enhanced input. In the target texts, there were a total of 46 target sentences out of 73 sentences in 890-word-long passages. Roughly half of the RCs contained the marker *who* and half *which*. Also, about half were embedded in the subject position and half in the object position in the matrix clause. As in most other studies on English relativization, this study used the preposition-stranding construction (e.g., *the student who(m) the teacher is looking for*) rather than the pied-piping one (e.g., *the student for whom the teacher is looking*).

4. The reliability of all tests was checked using the Kuder-Richardson 20 (K-R20) and Kuder-Richardson 21 (K-R21) reliability measures. In all cases, the values obtained were greater than .80 for K-R20 and greater than .78 for K-R21, which were taken to be reasonably high.

5. It might be useful to compare these effect sizes with those obtained in a similar study by Doughty (1988, 1991), from which the present study obtained many insights. The effect sizes of the meaning-oriented group and the rule-oriented group in her study were 1.34 and 1.16, respectively. Although we need to take into account various differences in the treatments used in the two studies (notably the amount of time for input exposure, which is substantially larger in Doughty's study; see Izumi, 2000, chap. 7 for a detailed comparison), their comparison underscores the effectiveness of the present output–input treatment.

6. The different results obtained for the first and second sets of univariate analyses may be explained by the fact that subdividing the output variable by the input enhancement variable resulted in a smaller n for each group, which led to reduction in the statistical power.

7. A possible explanation for the failure of the GJT to show any between-groups discrimination may be due to a ceiling effect because all groups started out with already relatively high scores on this test at the time of the pretest. The failure of the PCSCT to discriminate among groups and over time seems related to the general failure of this test to elicit the target RCs, despite the efforts to improve this test through piloting.

8. The data from four subjects were eliminated due to their failure to follow the given directions. The resultant N for this analysis was 43. The reliability of the scoring was checked by having two raters evaluate half of the recall summaries independently. An interrater correlation of $r = .84$ was obtained, which was taken to be reasonably high for the measure.

9. It is relevant to note that later work in cognitive psychology (e.g., Graf & Mandler, 1984) found that, although the level-of-processing effect is quite robust on explicit memory tests (e.g., cued recall test), it was not evident on implicit memory tests that do not require conscious effort to recollect specific prior episodes (e.g., stem completion test). In this regard, it is of interest to note that instructional effects in SLA studies are often measured using instruments that call on more explicit memory-based performance (Norris & Ortega, 2000). This study is no exception, and it is possible that different outcomes would have been obtained if more implicit types of assessment instruments had been used. Proper investigation of this issue would require careful consideration of at least three issues: (a) What type of tests constitutes implicit memory tests in SLA? (b) How applicable to SLA are the findings of adult subjects' showing automatic priming effects with familiar L1 words in implicit memory tests? and (c) What is the practical value of such an automatic activation effect if any such effect is obtained in SLA at all?

REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (Tech. Rep. No. 9, pp. 259–302). Honolulu: University of Hawai'i Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 77–117). New York: Cambridge University Press.

Brett, P. (1994). Using text reconstruction software. *ELT Journal, 48*, 329–336.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Doughty, C. (1988). *The effect of instruction on the acquisition of relativization in English as a second language*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Doughty, C. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition, 13*, 431–469.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). New York: Cambridge University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.),

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Focus on form in classroom second language acquisition (pp. 114–138). New York: Cambridge University Press.
Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.
Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
Farris, M. (1993). Libra: An authoring environment for multimedia lessons on the Macintosh (Version 1.1) [Computer software]. San Marcos, TX: Division of Media Services, Southwest Texas State University.
Gass, S. (1982). From theory to practice. In M. Hynes & W. Rutherford (Eds.), *On TESOL '81: Selected papers from the 15th annual conference of Teachers of English to Speakers of Other Languages* (pp. 129–139). Washington, DC: TESOL.
Graf, P. (1994). Explicit and implicit memory: A decade of research. *Attention and Performance, 15*, 681–696.
Graf, P., & Mandler, G. (1984). Activation makes words more accessible but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553–568.
Graf, P., & Schacter, D. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 930–940.
Hyltenstam, K. (1984). The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In R. Andersen (Ed.), *Second languages: A cross-linguistic perspective* (pp. 39–58). Rowley, MA: Newbury House.
Izumi, S. (2000). *Promoting noticing and SLA: An empirical study of the effects of output and input enhancement on ESL relativization*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.
Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34*, 239–278.
Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition, 21*, 421–452.
Jourdenais, R. (1998). *The effects of textual enhancement on the acquisition of the Spanish preterit and imperfect*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.
Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (Tech. Rep. No. 9, pp. 183–216). Honolulu: University of Hawai'i Press.
Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry, 8*, 63–99.
Kormos, J. (1999). Monitoring and self-repair in L2. *Language Learning, 49*, 303–342.
Kramer, S., & Rosenthal, R. (1999). Effect sizes and significance levels in small-sample research. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 59–79). Thousand Oaks, CA: Sage Publications.
Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.
Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.
Leow, R. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning, 8*, 151–182.
Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: Benjamins.
Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal, 82*, 357–371.
Long, M., & Robinson, P. (1998). *Focus on form: Theory, research, and practice*. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). New York: Cambridge University Press.
MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Mahwah, NJ: Erlbaum.
Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*, 338–356.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Nelson, K. (1987). Some observations from the perspective of the rare event cognitive comparison theory of language acquisition. In K. E. Nelson & A. van Kleek (Eds.), *Children's language* (Vol. 6, pp. 289–331). Mahwah, NJ: Erlbaum.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 13*, 417–528.
Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*, 186–214.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*, 283–331.
Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223–247.
Roth, F. (1984). Accelerating language learning in young children. *Journal of Child Language, 11*, 89–107.
Rutherford, W., & Sharwood Smith, M. (1985). Consciousness-raising and Universal Grammar. *Applied Linguistics, 6*, 274–282.
Saxton, M. (1997a). The contrast theory of negative input. *Journal of Child Language, 24*, 139–161.
Saxton, M. (1997b). *Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech*. Unpublished manuscript, University of London.
Schachter, J. (1998). Recent research in language learning studies: Promises and problems. *Language Learning, 48*, 557–583.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 206–226.
Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i Press.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237–326). Rowley, MA: Newbury House.
Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15*, 165–179.
Shook, D. (1994). FL/L2 reading, grammatical information, and the input-to-intake phenomenon. *Applied Language Learning, 5*, 57–93.
Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375–405). San Diego, CA: Academic Press.
Spada, N., & Lightbown, P. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *Modern Language Journal, 83*, 1–22.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review, 50*, 158–164.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seildlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.
Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). New York: Cambridge University Press.
Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371–391.
Tavakolian, S. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian (Ed.), *Language acquisition and linguistic theory* (pp. 167–187). Cambridge, MA: MIT Press.
Thornbury, S. (1997). Reformulation and reconstruction: Tasks that promote "noticing." *ELT Journal, 5*, 326–335.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 183–203.
Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition, 15*, 181–203.
Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*, 103–135.
VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Westport, CT: Ablex.
White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). New York: Cambridge University Press.
Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (pp. 71–105). New York: Wiley and Sons.
Williams, J. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition, 21*, 1–48.

https://doi.org/10.1017/S0272263102004023 Published online by Cambridge University Press