*SSLA*, **28**, 31–65. Printed in the United States of America.
DOI: 10.1017/S0272263106060025

# PROCESSING INSTRUCTION AND MEANINGFUL OUTPUT-BASED INSTRUCTION:

## *Effects on Second Language Development*

Kara Morgan-Short and Harriet Wood Bowden
*Georgetown University*

This study investigates the effects of meaningful input- and output-based practice on SLA. First-semester Spanish students ($n$ = 45) were assigned to processing instruction, meaningful output-based instruction, or control groups. Experimental groups received the same input in instruction but received meaningful practice that was input or output based. Both experimental groups showed significant gains on immediate and delayed interpretation and production tasks. Repeated-measures analyses of variance showed that overall, for interpretation, both experimental groups outperformed the control group. For production, only the meaningful output-based group outperformed the control group. These results suggest that not only input-based but also output-based instruction can lead to linguistic development.

It seems to be universally accepted that SLA is dependent on input (Gass, 1997; VanPatten, 2004a, 2004b). A significant amount of research has focused on one particular type of input-based instruction: processing instruction (PI). This research has supported the efficacy of PI (Benati, 2001, 2005; Cadierno,

We thank Ron Leow, Alison Mackey, and Cristina Sanz for their continual support and valuable input throughout the various phases of this research and for their comments on earlier versions of this manuscript. Likewise, we express our gratitude to Bill VanPatten and the various anonymous reviewers for their comments on the manuscript. We also thank the Linguistics Department for the use of its laboratory and equipment during data collection and Ru San Chen for assistance with statistical analyses. Special thanks to Gorky Cruz and Cristina Sanz for the use of digital photographs. Any errors or omissions are ours alone.

Address correspondence to: Kara Morgan-Short, Georgetown University, Department of Spanish & Portuguese, P.O. Box 571039, Washington, DC 20057-1039; e-mail: morgankd@georgetown.edu; or Harriet Wood Bowden, Georgetown University, Department of Spanish & Portuguese, P. O. Box 571039, Washington, DC 20057-1039; e-mail: woodh@georgetown.edu.

© 2006 Cambridge University Press 0272-2631/06 $12.00


https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

1995; Cheng, 2002; Farley, 2001a, 2001b; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Oikkenon, 1996; VanPatten & Sanz, 1995). In addition to offering evidence as to the effectiveness of PI, some researchers have made claims that PI directly affects the acquisition[1] of target forms, whereas output-based traditional instruction (TI)—to which PI has often been compared—results in learned linguistic knowledge (VanPatten & Cadierno). Although VanPatten (2002, p. 762) recently asserted that "output may play a number of important roles in language development" and that, in fact, it plays a facilitative role in acquisition (VanPatten, 2004b), he did not agree with the claim that "using a form in one's output is a direct path to acquisition" (VanPatten, 2004b, p. 27) and suggested instead, on the basis of current evidence, that acquisition does not appear to be dependent on output (VanPatten, 2004a).

There are at least two perspectives, however, that ascribe a more fundamental role to output in SLA. One perspective has suggested that both input and output develop corresponding comprehension and production skills (DeKeyser, 1997, 2001; DeKeyser & Sokalski, 1996). From a second perspective, the output hypothesis (Swain, 1993, 1995, 1998; Swain & Lapkin, 1995) holds that although input is essential to SLA, output might also bring about mental processes that both directly and indirectly affect acquisition. Research motivated by this hypothesis has supported a positive role for output practice in conjunction with input (e.g., Izumi, 2002; Izumi & Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Fearnow, 1999; Swain, 1995; Swain & Lapkin). Following the idea that both input- and output-based instruction can be effective for SLA, many studies have attempted to compare the two under a variety of research designs (Allen, 2000; Collentine, 1998; DeKeyser & Sokalski; Erlam, 2003; Nagata, 1998; Salaberry, 1997).

Few studies have specifically addressed whether output-based instruction can be as effective as PI when it is matched to PI on all variables other than practice mode. Recent studies by Benati (2005) and Farley (2001a, 2001b) have examined this question. Insofar as there was no difference between groups on production measures, these studies have been consistent; however, results have been somewhat inconsistent with regard to interpretation scores. In Farley (2001a) and Benati (2005), the PI groups were superior to output-based groups for interpretation, whereas in Farley (2001b), both experimental groups performed equally. Further research is called for to elucidate the efficacy of output-based instruction as compared to PI. Moreover, such additional research might shed light on the theoretical issue of the status of output within the broader scope of SLA.

## REVIEW OF LITERATURE

### Input and Processing Instruction Studies

Processing instruction is an input-based instructional technique informed by VanPatten's work on learners' input processing strategies (summarized in Van-

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

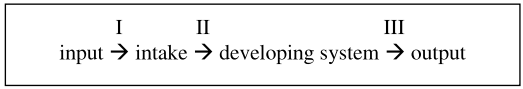



**Figure 1.** A sketch of basic processes in acquisition. From VanPatten and Cadierno, 1993a, p. 226.

Patten, 1996, 2004b). The theoretical model in which he represents the processes involved in SLA is presented in Figure 1, where I represents input processing, II represents accommodation and restructuring processes, and III represents processes involved in language production, such as monitoring, access, and control. PI is aimed at affecting learners' input processing and, thereby, should directly affect acquisition. To push learners to engage in more effective input processing, PI provides (a) explicit, nonparadigmatic grammatical instruction (i.e., forms and relevant examples presented sequentially) that includes input through examples and information about a processing strategy, (b) structured input practice composed of meaningful activities (both aural and written, and referential and affective), and (c) feedback. Thus, PI provides learners with two types of input (we consider the term *input* to refer to any type of second language [L2] linguistic data): input through examples and structured input. Structured input is "input that is manipulated in particular ways so that learners become dependent on form and structure to get meaning and/or to privilege the form or structure in the input so that learners have a better chance of attending to it" (VanPatten, 2002, pp. 764–765).

In the first PI studies, VanPatten and Cadierno (1993a, 1993b) empirically compared the effectiveness of PI and TI on the development of Spanish preverbal direct object pronouns by second-year Spanish students. TI presented learners with (a) paradigmatic grammatical instruction with input through examples and (b) oral and written output-based practice that moved from mechanical to meaningful to more open-ended communicative activities. Thus, PI and TI differed in both explanation (nonparadigmatic, with processing strategy information, vs. paradigmatic, with no such information) and practice (always meaningful and input-focused vs. sometimes meaningful and output-focused). Participants underwent PI, TI, or formed part of a control group that was not exposed to the target forms. Results from the two studies (see Table 1), which followed the same design, found that for interpretation, PI significantly outperformed TI and control.[2] For production, PI and TI were not significantly different from each other in either study. However, in VanPatten and Cadierno (1993b), only TI was significantly better than control for production. Overall, considering both interpretation and production results from both studies, PI appears to show an advantage over TI. VanPatten and Cadierno's interpretations of these findings, however, merit careful consideration.

First, VanPatten and Cadierno (1993b) argued that PI succeeded in affecting the developing system, whereas TI did "not feed into the developing system directly but instead result[ed] in a different knowledge system" (p. 53). This conclusion was on the basis of the findings that (a) for interpretation, PI

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press



**Table 1.** Summary of processing instruction studies

| Study | *N* | Participant level | Linguistic feature | Experimental groups | Assessment tasks[a] | Results |
|---|---|---|---|---|---|---|
| VanPatten and Cadierno (1993a) | 80 | 2nd year university Spanish | Spanish DO pronouns | PI<br>TI<br>C [−exposure][b] | Interpretation<br>Production<br>Sentence-level completion | PI > TI and C<br><br>PI and TI > C |
| VanPatten and Cadierno (1993b) | 49 | 2nd year university Spanish | Spanish DO pronouns | PI<br>TI<br>C [−exposure] | Interpretation<br>Production<br>Sentence-level completion | PI > TI and C<br><br>TI > C |
| Cadierno (1995) | 61 | 3rd semester university Spanish | Spanish past tense | PI<br>TI<br>C [−exposure] | Interpretation<br>Production<br>Sentence-level completion | PI > TI and C<br><br>PI and TI > C |
| VanPatten and Sanz (1995) | 44 | 3rd semester university Spanish | Spanish DO pronouns | PI<br>C [−exposure] | Interpretation<br>Production<br>Written and oral sentence-level completion<br>Video narration<br>Structured oral interview | PI > C<br><br>PI > C<br><br>PI = C<br>PI > C |
| Benati (2001) | 39 | 2nd semester university Italian | Italian future | PI<br>output<br>C [+exposure] | Interpretation<br>Production<br>Written verb conjugation<br>Oral sentence-level production | PI > output > C<br><br>PI and output > C<br>PI and output > C |
| Farley (2001a) | 29 | 4th semester university Spanish[c] | Spanish subjunctive | PI<br>MOI | Interpretation<br>Production<br>Sentence-level completion | PI > MOI<br><br>PI = MOI |

| | | | | | | |
|---|---|---|---|---|---|---|
| Farley (2001b) | 50 | 4th semester university Spanish | Spanish subjunctive | PI<br>MOI | Interpretation<br>Production<br>Sentence-level completion | PI = MOI<br><br>PI = MOI |
| Cheng (2002) | 109 | 4th semester university Spanish | Spanish copulas *ser* and *estar* | TI<br>PI<br>C [−exposure] | *Ser* and *estar* combined<br>Interpretation<br><br>Production<br>Sentence-level completion<br><br>Composition<br><br>*Estar* only<br>Interpretation<br><br>Production<br>Sentence completion<br><br>Composition<br><br> | <br>Post 1: PI > C<br>Post 2: TI > C<br><br>Post 1: PI > C<br>Post 2: PI and TI > C<br>Post 1: PI and TI > C<br>Post 2: PI and TI > C<br><br>Post 1: PI > C<br>Post 2: PI = TI = C<br><br>Post 1: PI and TI > C<br>Post 2: PI > C<br>Post 1: PI and TI > C<br>Post 2: PI = TI = C |
| Benati (2005) | 77 | Secondary school English[c] | English simple past tense | PI<br>MOI<br>TI | Interpretation<br>Production<br>Sentence-level completion | PI > MOI and TI<br><br>PI = MOI = TI |

*Note*. PI = Processing instruction; TI = Traditional instruction; C = Control group; MOI = Meaning-based output instruction; DO = Direct object. Significant differences are noted by the > sign. Nonsignificant differences are noted by the = sign.

[a]Interpretation tasks are aural and production tasks are written unless otherwise noted.

[b][± exposure] indicates whether the control group was exposed to the target form.

[c]Study indicated that participants had no prior formal exposure to target linguistic form at the institution indicated.



https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

outperformed TI and the control group and (b) for production, there were no significant differences between PI and TI. VanPatten and Cadierno deemed their interpretation of these results to be consistent with Krashen's (1982) learning/acquisition distinction. Although this explanation is plausible, it is not the only possible explanation (e.g., both instructional techniques might have resulted in one type of knowledge, with different degrees, amounts, or strengths of knowledge) and, in fact, their claim might be empirically unfalsifiable (Salaberry, 1997). Note that in more recent work, VanPatten (2002) has focused on the relative effectiveness of PI over TI and has neither reiterated nor denied the claim that PI leads to acquisition but TI leads to a different knowledge system.

Second, the authors concluded that their results supported Schwartz's (1993) theoretical claim that the language module can operate only on primary linguistic data (input) and that explicit data and negative evidence cannot be used by the language module. Although their results are indeed consistent with this view, the study did not isolate structured input as a variable; instead, the PI group received both explicit data and negative evidence in addition to structured input. Therefore, as pointed out by Salaberry (1997) and by an anonymous *SSLA* reviewer, it is impossible to determine whether the effectiveness of PI is due to input alone or to input in combination with explicit data and negative evidence.

A third issue relates to the role of output practice in SLA. VanPatten and Cadierno (1993a) concluded that "instruction is apparently more beneficial when it is directed at how learners perceive and process input rather than when it is focused on practice via output" (p. 240). Although the authors advocated the use of output activities in instruction for the development of fluent and accurate production, they also stated that "the value of grammatical instruction as output practice is questionable if the attempt of the instruction is to alter the nature of the developing system" (VanPatten & Cadierno, 1993b, p. 46). These studies, however, were not designed to isolate mode as a variable, as pointed out by VanPatten (2002) and DeKeyser, Salaberry, Robinson, and Harrington (2002). In fact, VanPatten emphasized that "processing instruction cannot be reduced to input/comprehension and that traditional instruction cannot be reduced to output/production" (p. 827), although these issues have often been conflated. Therefore, given that the variable of mode was not isolated, the authors' conclusions regarding output practice seem to be overly strong. Moreover, these conclusions might have contributed to the ongoing interest in examining the role of output in relation to input.

Subsequent to VanPatten and Cadierno (1993a, 1993b), VanPatten and other researchers undertook replications of the initial studies, employing new target forms and a wider variety of task measures (see Table 1). Cadierno (1995) compared the effects of PI and TI on the development of a different target form: the Spanish preterit tense. Her results directly paralleled those of VanPatten and Cadierno (1993a). VanPatten and Sanz (1995) sought to extend the generalizability of the original studies to tasks above the sentence level. Although learners' performance varied on different tasks, the observed ben-

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

efits of PI (here compared only to a control group) held for suprasentential tasks in both oral and written modes, with the exception of an oral video narration task (arguably the most cognitively demanding of the tasks).

Cheng (2002) compared PI and TI on a different target form—the Spanish copulas *ser* and *estar*—and with additional task measures. Data from *ser* and *estar* did not replicate findings from VanPatten and Cadierno (1993a). Subsequent analyses of data related to *estar* only showed that PI outperformed control more often than TI and that, again, PI and TI were never statistically different. On the basis of this set of data, Cheng claimed that "PI appears more effective in helping students make correct form-meaning mapping [*sic*] and in restructuring their mental representation of target forms" (p. 317). This claim has been controversial in the literature.[3] Overall, given that there were no significant differences between PI and TI on any posttest under either combined or separate analysis, the claim that PI was more effective than TI in restructuring learners' mental representations does not appear to be strongly supported. Cheng suggested one further explanation for the similar performance of her PI and TI groups: Upon further review of her TI instructional materials, she found that they were "substantially more meaningful and communicative" (p. 316) than those of VanPatten and Cadierno (1993a). The increased similarity in meaningfulness between Cheng's PI and TI treatment conditions coupled with the lack of difference in performance for these groups suggest that meaningfulness might be an important variable to consider.

To address the issue of meaningfulness, Benati (2001) compared PI to a more meaningful output instruction on the Italian future tense. The results mirrored those of VanPatten and Cadierno (1993a) in that, for interpretation, PI outperformed the output and control groups, and for production, there was no statistical difference between instructional treatments. The findings were unique, however, in that the output group also outperformed the control group on interpretation. This difference was likely due to the added meaningfulness of his output condition, as acknowledged by VanPatten (2002). Thus, the variable of meaningfulness appears to play a role in the effectiveness of instruction, an issue we will return to later.

In sum, the research on PI indicates that (a) PI leads to linguistic development for different forms, as evidenced by improved performance on various tasks (Benati, 2001; Cadierno, 1995; Cheng, 2002; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz, 1995), and (b) PI is more effective than TI in some cases (Cadierno, 1995; VanPatten & Cadierno, 1993a, 1993b); although when TI is more meaningful or when the target form is more semantically complex—as pointed out by DeKeyser et al. (2002)—the advantage of PI is not as clear (Benati; Cheng). These studies are not without methodological concerns, however, which will be considered in a later section. Moreover, it remains unclear whether these studies offer evidence that supports claims made by Krashen (1982) and Schwartz (1993) with regard to the acquisition/learning distinction and whether only primary linguistic data affect the developing system. Additionally, it appears that the meaningfulness of practice might have some effect

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

on linguistic development (Benati; Cheng) and, thus, should be explored further. Finally, the role of output—although not a principal goal of investigation of these studies—is clearly of interest and remains to be clarified.

## The Output Hypothesis and Relevant Studies

Although it has not been within the scope of most PI research to clarify what role(s) output might play in SLA, some researchers (e.g., Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999; Swain, 1995; Swain & Lapkin, 1995) have specifically set out to investigate this issue. These researchers, working from an output-based perspective, do not deny the essential role of input in SLA. They do, however, reject the view (e.g., Krashen, 1989; Schwartz, 1993) that input alone can directly affect the developing linguistic system (Izumi; Swain, 1995). The output hypothesis (Swain, 1993, 1995, 1998; Swain & Lapkin) suggests that output might bring about mental processes that are part of the process of L2 learning and posits three functions through which output might play an active role in L2 learning: the hypothesis-testing function, the metalinguistic function, and the noticing function.

Evidence from research that supports some of the functions of the output hypothesis suggests that output might indeed be beneficial for SLA (Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999; Swain, 1998; Swain & Lapkin, 1995). Izumi et al. and Izumi and Bigelow compared an experimental group, which received input via written exposure to the target form and engaged in output tasks, to a comparison group, which received the same exposure to the target form but did not engage in output. With only one exception, in which the experimental group outperformed the comparison group, there were no statistical differences between groups on any measure. The general lack of difference between groups was attributed to task demands rather than the learning conditions. Following a similar study design, but one that reduced task demands, Izumi compared four experimental groups (composed of combinations of ± output and ± input enhancement) and a control group. The +output groups engaged in a text reconstruction task, whereas the −output groups answered extension questions based on the text. Results indicated that participants in the +output groups used the target form in the reconstruction tasks and outperformed nonoutput and control groups on posttest measures.

In sum, empirical evidence from two studies investigating the output hypothesis (Izumi, 2002; Izumi et al., 1999) lends at least some support to the notion that output might have beneficial effects on linguistic development in addition to—not in opposition to—the crucial role of input. These studies, taken in conjunction with more qualitative work by Swain and colleagues (Swain, 1998; Swain & Lapkin, 1995), motivate further research into the role of output in SLA. An interpretation of this line of research is that output might affect acquisition, not just learned linguistic knowledge. However, as in the case of PI research, a conclusion regarding this distinction is, as yet, unwarranted.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

## Studies Examining Relative Effects of Input and Output

Another strand of research (Allen, 2000; Collentine, 1998; DeKeyser & Sokalski, 1996; Erlam, 2003; Nagata, 1998; Salaberry, 1997) focused on the exploration of any differential effects of input-based as compared to output-based instructional conditions. Accordingly, much of this research has made attempts to methodologically isolate the variable of output. Specifically, these studies have compared the effects of some type of input practice (often labeled PI) to an output practice (see Table 2). For instance, in order to isolate the variable of practice mode, DeKeyser and Sokalski held grammatical explanation constant and found that the input group outperformed the control group on a comprehension task, whereas the output group outperformed control on a production task. This result was viewed as consistent with skill acquisition theory, and the researchers concluded that "comprehension and production skills in an L2 are to some extent learned separately" (DeKeyser & Sokalski, p. 635) and interpreted the results as evidence contrary to those of VanPatten and Cadierno (1993a, 1993b). Results from Salaberry, Nagata, Allen, and Erlam have also been interpreted as contrary to those of the PI studies mentioned earlier in the literature review. Salaberry reported that both input and output groups outperformed control on the comprehension task, but neither was better than control on production tasks. Nagata's results indicated that both input-based and output-based practice were equally effective for comprehension tasks, but that output-based practice was superior for production tasks. Allen found that both PI and TI performed similarly on an interpretation task, whereas TI outperformed PI on a production task. Erlam found that an output-based group outperformed a structured-input group on a listening comprehension test and on form placement in a written production test; however, no significant differences were found between groups on a reading comprehension test, on form correctness in the written production test, or on an oral production test.

The results from these studies contribute to the field's understanding of how input and output affect learners' comprehension and production of L2 target forms and structures. Care should be taken, however, when comparing these studies' results to those of VanPatten and colleagues' studies on PI and TI, given that the treatments and assessment tasks are not equivalent across the studies (see Sanz & VanPatten, 1998; VanPatten, 2002; VanPatten & Wong, 2004, for additional discussion). Treatments are substantially different from those in PI research in that (a) they did not provide explicit information designed to alter any learner processing strategy (DeKeyser & Sokalski, 1996; Erlam, 2003; Nagata 1998), (b) they (either intentionally or unintentionally) reduced the meaningfulness of input activities (DeKeyser and Sokalski; Erlam[4]; Salaberry, 1997), (c) they did not control event probabilities (Allen, 2000), and (d) the output group performed some interpretation activities during treatment (Allen). Assessments and scoring procedures used in these studies also differed from those of PI studies in that (a) comprehension instead of interpretation tests were given (see Note 2) (DeKeyser & Sokalski; Erlam; Nagata;

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press



**Table 2.** Summary of studies examining relative effects of input and output

| Study | *N* | Participant level | Linguistic feature | Experimental groups[a] | Assessment Tasks[a,b] | Results |
|---|---|---|---|---|---|---|
| DeKeyser and Sokalski (1996) | 36 | Spanish I university | Spanish DO pronouns | IP | Comprehension | IP > C |
| | | | | OP | Production | |
| | | | | C [+ 10 min. explanation] | Translation | OP > C |
| | | | | | Fill-in-the-blank | OP > C |
| | 46 | Spanish II university | Spanish conditional | | Comprehension | OP > C |
| | | | | | Production | IP and OP > C |
| Salaberry (1997) | 26 | 3rd semester university Spanish | Spanish DO pronouns | IP | Comprehension (of written input) | IP and OP > C |
| | | | | OP | Production | |
| | | | | C [−exposure][c] | Sentence-level completion | IP = OP = C |
| | | | | | Free narration | IP = OP = C |
| Collentine (1998) | 54 | 2nd semester university Spanish | Spanish subjunctive | PI | Interpretation | PI and OP > C |
| | | | | OP | Production | |
| | | | | C [−exposure] | Fill-in-the-blank | PI and OP > C |
| Nagata (1998) | 14 | 2nd semester university Japanese[d] | Japanese honorifics | IP | Comprehension (5 tasks) | OP = IP |
| | | | | OP | Production | |
| | | | | | 5 matched tasks | OP > IP |
| | | | | | Oral conversation | OP > IP |
| Allen (2000) | 179 | 4th semester high school French[d] | French causative | PI | Interpretation | PI and TI > C |
| | | | | TI | Production | |
| | | | | C [−exposure] | Sentence generation | TI > PI > C |
| Erlam (2003) | 70 | 2nd year secondary school French[d] | French DO pronouns | IP | Comprehension | |
| | | | | OP | Reading | OP > C |
| | | | | C [−exposure] | Listening | OP > IP and C |
| | | | | | Production | |
| | | | | | Written | IP and OP > C (placement & form) |
| | | | | | | OP > IP (placement only) |
| | | | | | Oral | OP > C (# attempts) |

*Note*. PI = Processing instruction; TI = Traditional instruction; C = Control group; IP = Input practice; OP = Output practice; DO = Direct object. Significant differences are noted by the > sign. Nonsignificant differences are noted by the = sign.

[a]Terminology follows the original studies, but see Note 2.

[b]Interpretation and comprehension tasks are aural and production tasks are written unless otherwise noted.

[c][± exposure] indicates whether the control group was exposed to the target form.

[d]Study indicated that participants had no prior formal exposure to target linguistic form at the university level.

Salaberry), (b) no sentence-level production test was included (Allen), and (c) production tests were either scored so as not to give credit for partially correct answers (DeKeyser & Sokalski; Salaberry) or only particular morphological aspects of the target form were considered for scoring (Erlam). The differences in treatments and assessments used in these studies make it difficult to directly compare their results to those of studies designed under the PI framework.

Collentine's (1998) study maintained more treatment fidelity to PI itself but still found results that appear to contradict those of the original studies. This study compared PI to an output-oriented instruction for the Spanish subjunctive. The results revealed no significant differences between experimental groups on either comprehension or production tasks; both were superior to control. Collentine concluded that the similarity in performance between groups was due to the high communicative value of tasks in both conditions, which might have compelled both groups of learners to make form-meaning connections irrespective of practice mode. Although some aspects of Collentine's PI treatment have been criticized (see Farley, 2001b, 2002; VanPatten, 2002), the treatment appears to maintain sufficient fidelity for comparison (see Collentine, 2002; Farley, 2004). Therefore, Collentine's finding of no differences between the instructional groups offers some evidence that PI is not necessarily superior to all output-based instruction.

In sum, the majority of studies that have investigated the relative effects of input and output treatments are difficult to compare to the PI research of VanPatten and colleagues because their input treatments, or assessment tasks, or both were not equivalent to those used in PI research. These studies might best be interpreted not as contradictory to PI studies, but as addressing a different but related issue. Broadly speaking, studies from this strand of research suggest that output-based instruction is at least as effective as input-based instruction (Allen, 2000; Collentine, 1998; DeKeyser & Sokalski, 1996; Erlam, 2003; Nagata, 1998; Salaberry, 1997). Only three of these studies (DeKeyser & Sokalski; Nagata; Salaberry), however, were designed to methodologically isolate the variable of practice mode (i.e., input or output) with attempts made to control all other variables (e.g., grammatical explanation, meaningfulness of practice). Indeed, these three studies did find that output-based practice was at least as effective as input-based practice for linguistic development. This type of methodology is essential if the goal is to compare the two. Further, if one aims to study this contrast within the framework of PI, the input-based treatment must be faithful to the tenets of PI and the output-based treatment must be matched to PI on all variables other than mode.

## Meaning-Based Output Instruction Studies

Recent studies by Farley (2001a, 2001b) and Benati (2005) successfully compared PI to meaning-based output instruction (MOI) on the development of

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

the Spanish subjunctive, by controlling all variables other than mode and maintaining treatment fidelity to PI. In Farley (2001a, 2001b), MOI was very similar to PI in that it included nonparadigmatic grammatical explanation, information on processing strategy, and input through examples; however, its practice was output based rather than input based. Both conditions provided learners with feedback at the end of each practice activity. For PI, it is unclear whether this feedback included the target form (e.g., whether the correct answer was repeated or merely identified as option A or B). For MOI, the instructor solicited answers from participants until the correct answer was given. Both of Farley's studies followed a similar design and found that the PI and MOI groups made gains on both interpretation and production measures, although no control group was included for comparison. However, the results from Farley's two studies differed: Whereas Farley (2001a) found that (a) PI was superior to MOI on the interpretation task and (b) there were no differences between groups on the production task (results that were similar to previous PI studies), Farley (2001b) found no differences between PI and MOI on either task. Farley (2001b) proposed that the MOI group might have been able to perform as well as the PI group on the interpretation task because the review of answers in the MOI group might have been "utilized as focused input for other learners" (p. 81). In other words, this review might have served as incidental input. As to why these results differed from his previous (2001a) study, Farley offered two potential explanations. First, in the first study (2001a), participants completed 8 activities, whereas in the second study (2001b), they completed 10. Therefore, both groups would have received more practice in the second study, which included, as a result, more incidental input in the MOI condition. Second, participants in the first study (2001a) were products of a regular track of university-level Spanish study, whereas those in the second study (2001b) were products of a more intensive track of study; thus, their prior experience differed. Even though these explanations could potentially account for higher scores overall for the second study (2001b), there is no underlying theoretical justification for why these factors would have affected the treatment groups differentially; that is, they fail to explain why the MOI group would have benefited more from these factors than the PI group.

A more recent study (Benati, 2005) compared the effectiveness of PI, TI, and MOI on linguistic development of the English simple past tense for secondary school participants (ages 12–13 years). PI and MOI were matched on all variables except for mode of practice, whereas TI differed in that (a) it did not provide information on processing strategy and (b) only half of the practice provided was meaningful. For all groups, feedback indicated only whether learners' answers were correct or incorrect, which for MOI would imply that the participants supplied their answers aloud in the class and then received feedback, as in Farley (2001a, 2001b). To assess linguistic development, Benati administered pretests and posttests, but not delayed tests. Analyses indicated that all groups improved on both interpretation and production. Furthermore, for interpretation, PI outperformed both TI and MOI, whereas for

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

production, all groups performed similarly. Benati concluded that PI has a positive effect on learners' developing systems and suggested that although output plays a possible role in language acquisition, it is "not responsible for the making of an implicit system" (p. 85), a conclusion that is similar to those drawn by VanPatten and Cadierno (1993a, 1993b) and VanPatten (2004b). Given the nonconverging results of Farley's and Benati's studies, it is important for further research to be carried out in order to determine whether PI and MOI have differential effects.

## Theoretical and Methodological Motivations for Current Study

The motivation for this study was (a) to investigate whether meaningful output-based instruction (MOBI) and practice could, in conjunction with input, lead to linguistic development and (b) to compare this linguistic development, if any, to that of PI for Spanish direct object pronouns. More broadly, such a study allows for exploration of the role of output in SLA in a meaningful context and under the PI framework. The current study maintained strict treatment fidelity to PI: The grammatical explanation, PI treatment, and assessment tasks were adapted directly from those used by VanPatten and Cadierno (1993a, 1993b), and the target form was the same. MOBI was matched to PI in all respects except practice mode. This research design was similar to that of Farley (2001a, 2001b) and Benati (2005), but differed in that it examined the same target form as that of VanPatten and Cadierno (1993a, 1993b) and was computer based, which allowed for greater methodological control.

Along these same lines, we endeavored to address four specific limitations of previous studies discussed in this literature review. First, approximately one third (4 of 15) did not include a control group in their research design. Of those that did, the control group was most often not exposed to the target forms or structures (9 of 11). It seems that these control groups were employed to address potential learning due only to test effects. Thus, the benefits of experimental conditions as compared to simple exposure to target forms could not be assessed. Most previous research may not have controlled for participants' prior knowledge. Few studies (5 of 15) indicated that learners had no prior formal exposure to the target structure and, thus, it is possible that treatments in the majority of the studies interacted with or triggered learners' prior knowledge, creating an uncontrolled moderator variable that would not be present in naïve learners. Further, many studies used seemingly high pretest cutoff points for participant selection (60–80%). Thus, it was likely that many participants indeed had some prior knowledge of the target form and, therefore, an interaction between prior knowledge and instruction cannot be ruled out when interpreting results. Third, the element of feedback was not always carefully reported in many of the studies. As many were classroom based, feedback was naturally provided to the class as a whole rather than to the individual, and it was not always stated whether this feedback contained the

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

target form or merely referred to the letter or number of the correct choice. For output-based treatments, answers were provided either by students or teachers and thus necessarily included the target form. Therefore, all participants in output-based treatments received target-form input for each practice item, regardless of whether they needed correction on the item. A final possible intervening variable that has often gone unaddressed is that of experimenter bias or differences between instructors when more than one was used. The design of the current study was specifically developed to better address these methodological limitations in the following ways: (a) The control group was exposed to the target form and was approximately matched for time on task; (b) an attempt to control prior knowledge of the target form was made both by selecting first-semester Spanish students as participants and by using a pretest with a stricter criterion for inclusion ($< 33\%$), which was chance level for the interpretation task; (c) feedback was individualized through the computer-based treatment and was provided to participants in the output-based group only when they did not produce the correct answer; and (d) computer delivery of instruction eliminated potential differences between instructors and experimenter bias. Thus, in addition to matching treatments on all variables other than practice mode, the study's design addressed several additional methodological limitations of prior research.

The current laboratory study followed a pretest, treatment, immediate posttest, delayed posttest design. The dependent variable was linguistic development, as measured by tests of participants' interpretation and production of Spanish preverbal direct object pronouns. The independent variable was instructional type (PI and MOBI, which were compared to control). The research questions that motivated the study were the following:

1. Do PI and MOBI lead to improved performance on interpretation and production of Spanish direct object pronouns?
2. If both instructional conditions lead to improved performance, is the performance evidenced by learners in PI and MOBI groups comparable?

## METHOD

### Participants

The initial participant pool consisted of 51 students enrolled in a first-semester Spanish course at a northeastern university in the United States; the course was communicative in nature and used the *¿Sabías que. . . ?* course materials (VanPatten, Lee, & Ballman, 2000). Participants' ages ranged from 17 to 30 years, with an average of 19.3 years. Information on previous language study was collected.[5] Participants were assigned randomly to one of two instructional groups (PI or MOBI) or to a control group. First-semester students were selected in order to minimize prior knowledge of the target form. It should be noted that although participants had not yet been formally exposed to the target

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

form during the semester, the study was conducted immediately prior to the scheduled introduction of the target form in class and, thus, its presentation should have been appropriate at this point. No explicit instruction or assignments on the target form were given prior to or during the course of the study. In order to further control for prior knowledge of the target form, participants who scored 33% or greater (i.e., better than chance on interpretation) on either the interpretation or the production pretest were eliminated.[6] After eliminations due to pretest scores, the final subject pool consisted of 45 participants. Participant attrition was approximately equal in all groups, and a one-way ANOVA was performed to ensure that the three groups did not differ with regard to prior Romance language study; there were no significant differences between groups $F(2, 42) = 0.22$, $p = .80$.

## Target Structure

The target form used in this study was Spanish preverbal direct object pronouns, following several previous PI studies (Cadierno, 1995; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz, 1995). It has been established through child and adult studies (see VanPatten & Cadierno, 1993a, p. 47; VanPatten, 2004b for discussion) that native English-speaking L2 learners use the first-noun strategy when processing noun-verb-noun strings and, thus, are likely to incorrectly process pronouns as subjects when they are in sentence-initial position. Because this is the case with preverbal direct object pronouns, the target form is amenable to the tenets of PI, which require that there be a strategy that can be countered.

## Materials

The computer-delivered treatments were developed using Authorware 5. The PI treatment was adapted directly from the pen-and-paper instructional packet used by VanPatten and Cadierno (1993a, 1993b).[7] The present study investigated the effects of PI compared to MOBI, which was developed to match PI in its explanation and meaningfulness, but was contrasted to PI in the mode of practice provided. Thus, both experimental groups received identical grammatical explanation that was nonparadigmatic, included explicit information on the structure, gave information about the first-noun processing strategy, and provided input in the form of examples. Meaningfulness during practice activities was also held constant for both groups. Practice in each group included referential and affective activities, 3 aural or oral (for PI and MOBI, respectively) and 10 written. Thus, PI and MOBI were identical in structure and content; the fundamental difference was found in the mode of the practice activities, as described in the paragraph that follows.

In PI, practice was structured so that direct object pronouns had to be interpreted in order to complete the activities. No practice activity in this condition required the production of the target form (see Appendix A). MOBI

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

activities were adapted directly from the study's PI activities and modified to require production (rather than interpretation) of direct object pronouns in order to complete each activity (see Appendix B). Participants were never required to interpret the target form in order to complete practice activities, although there were two activities based on reading passages that included direct object pronouns. The control group was exposed to target forms via reading passages and comprehension questions that contained approximately equal numbers of the target forms as in the experimental conditions (see Appendix C). Specifically, participants were instructed to briefly study a vocabulary list for each passage (which remained accessible), to read the passage, and then to answer the questions that followed. Two of the four passages were taken from VanPatten and Cadierno (1993a, 1993b).

Feedback was designed to be comparable between PI and MOBI groups. In previous studies (Cadierno, 1995; Farley, 2001a, 2001b; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz, 1995), participants in PI groups received feedback that informed them of the correct answer for each exercise. In our study, to match experimental groups while remaining faithful to PI, the decision was made to provide this information (i.e., to identify the correct answer) to participants in both the PI and MOBI groups. For the PI group, all but two activities consisted of two choices. Feedback indicated whether the participant's response was correct or incorrect. After receiving such feedback, the correct answer should have been apparent to participants because the activity was still visible on the screen. The remaining two activities required interpretation of the target form in paragraph-level discourse. For these activities, the correct answer—the identification of the subject or object—was provided. For the MOBI group, feedback likewise indicated whether participants' responses were correct or incorrect. (The software considered only the minimal response required—the direct object pronoun or, in some exercises, the pronoun and the verb—and disregarded any surrounding context.) However, because activities in the MOBI group involved production, feedback that only indicated that a response was incorrect would not give participants access to the correct answer. For written activities, MOBI participants were therefore provided with the correct answer only when they responded incorrectly. For oral activities, because the computer program could not assess oral production, the correct answer was always provided (via written feedback). Unfortunately, the computer could not be programmed to track occurrences of incorrect answers or provision of feedback, so no data were available for analyses based on the provision of input via feedback for the MOBI group. For the control group, no feedback of any kind was provided.

## Assessment Tasks

Pretests, immediate posttests, and delayed posttests measured interpretation and production of the target structure. There were 20 items in the interpretation test, 12 of which were critical. Participants were given 15 s to listen to a

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

prerecorded sentence twice and to choose which of three pictures accurately depicted the sentence (see Appendix D). The three pictures depicted the same action, but the agent and patient of the verb varied. Therefore, a correct response hinged on correct identification of the agent and patient roles. Participants were also given the option to mark "I don't know" rather than being forced to make a guess. The eight distracter items contained no direct object pronouns and included either full noun phrases or reflexive pronouns. Likewise, the production task was composed of 20 items, of which 12 were critical. Participants were allowed 20 s to complete a sentence that corresponded to two related pictures using a verb that was provided (see Appendix E). The relationship between these pictures required the production of a direct object pronoun. The eight distracters required the use of a full noun phrase or a reflexive pronoun.

## Procedure

Approximately 2 weeks prior to the start of the experiment, participants completed vocabulary instruction and practice in order to ensure that they were familiar with vocabulary present in tests and treatments. The experiment was carried out in three sessions that spanned approximately 3 weeks. The language background questionnaire was given along with the pretests in class and, 2 weeks later, the computer-based treatments were administered individually (i.e., participants underwent either PI, MOBI, or control group activities).[8] Participants progressed through the treatments at their own pace. Hence, the precise duration of the treatment varied, with most participants completing the treatment in approximately 1 hr (no differences were noted among groups, although precise time on task for each participant was not recorded). The first posttest immediately followed the treatment, and the delayed posttest was administered 1 week later in class. A split-block design was followed, using three balanced versions of the tests. Participants completed an exit questionnaire after the delayed posttest, to control for external exposure to the target form.

## Scoring

Scoring procedures followed those of VanPatten and Cadierno (1993a, 1993b). The multiple-choice interpretation test was scored by awarding one point per item if the interpretation was correct and no points if incorrect. Production test items were worth a possible two points: Two points were awarded for correct form and placement of the direct object pronoun, one point was awarded for any other use of an object pronoun (correct form, wrong placement; incorrect form, correct placement; or incorrect form, incorrect placement), and no points were awarded if no direct object pronoun was used. This liberal scoring procedure was used, as in previous studies, in order to reveal

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

any partial effects of instruction. Scores were converted to percentages by dividing the number of points awarded by the total number of points possible.[9]

## RESULTS

Our first research question was: Do PI and MOBI lead to improved performance on interpretation and production of Spanish direct object pronouns? To answer this question, within-group repeated-measures ANOVAs were first carried out on interpretation scores. For interpretation, main effects for time were found for PI, MOBI, and control groups, $F(2, 28) = 54.99$, $p < .001$; $F(2, 26) = 26.89$, $p < .001$; and $F(2, 26) = 8.56$, $p < .01$, respectively. Given these significant results, individual one-way repeated-measures ANOVAs were carried out on interpretation data in order to analyze the change in scores between times for each group. The analyses revealed that both immediate posttest and delayed test scores for both PI and MOBI were significantly higher than pretest scores. Although the control group did not improve significantly from the pretest to the immediate posttest, a significant gain was made between the pretest and the delayed test. Analyses of the change in scores from immediate to delayed posttests revealed no significant change for PI, a significant loss for MOBI, and a significant gain for control. It must be kept in mind, however, that although MOBI evidenced a significant loss from the immediate to the delayed posttest, the delayed test scores were, nevertheless, significantly higher than the pretest scores for this group. See Tables 3 and 4.

Within-group repeated-measures ANOVAs were likewise carried out on production scores. Main effects for time were found for PI and MOBI, $F(2, 28) = 7.68$, $p < .01$ and $F(2, 26) = 20.36$, $p < .001$, respectively, but not for control, $F(2, 26) = 2.51$, $p = .10$. On the basis of these results, individual one-way repeated-measures ANOVAs were carried out to analyze change in scores on production measures at each time for PI and MOBI. As with interpretation, both immediate posttest and delayed test scores for both PI and MOBI were significantly higher than pretest scores. Analyses of the change in scores from immediate to delayed posttests revealed no significant change for PI and a significant loss for MOBI. Again, it should be noted that although MOBI evi-

**Table 3.** Mean percentage correct for interpretation tests

| | | Preinterpretation | | Postinterpretation | | Delayed interpretation | |
|---|---|---|---|---|---|---|---|
| Group | *N* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| PI | 15 | 7.62 | 7.04 | 67.38 | 23.17 | 62.22 | 25.56 |
| MOBI | 14 | 8.13 | 8.28 | 71.21 | 24.98 | 46.43 | 37.51 |
| Control | 14 | 9.96 | 9.99 | 13.69 | 18.95 | 35.71 | 29.31 |

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

**Table 4.** ANOVA results for time: interpretation tests

| Source of variability | *df* | Sums of squares | Mean squares | *F* | *p* | Partial eta$^2$ | Power |
|---|---|---|---|---|---|---|---|
| Pretest to immediate posttest | | | | | | | |
| PI | 1 | 28,565.28 | 28,565.28 | 83.69* | .00 | .85 | 1.00 |
| MOBI | 1 | 29,844.41 | 29,844.41 | 93.93* | .00 | .87 | 1.00 |
| Control | 1 | 97.65 | 97.65 | 0.48 | .50 | .036 | .10 |
| Pretest to delayed test | | | | | | | |
| PI | 1 | 22,855.01 | 22,855.01 | 64.29* | .00 | .82 | 1.00 |
| MOBI | 1 | 10,597.53 | 10,597.53 | 15.77* | .00 | .55 | .96 |
| Control | 1 | 4,644.53 | 4,644.53 | 10.71* | .01 | .45 | .86 |
| Immediate posttest to delayed test | | | | | | | |
| PI | 1 | 477.68 | 477.68 | 1.37 | .26 | .09 | .19 |
| MOBI | 1 | 3,806.99 | 3,806.99 | 7.12* | .02 | .35 | .69 |
| Control | 1 | 3,395.26 | 3,395.26 | 10.84* | .01 | .46 | .86 |

*$p < .05$.

denced a loss from posttest to delayed test, the delayed test production scores were significantly higher than the pretest production scores. See Tables 5 and 6. Thus, the results indicate that our first research question is answered affirmatively: Both PI and MOBI led to linguistic development for Spanish preverbal direct object pronouns, given that both immediate posttest and delayed test scores for both PI and MOBI were significantly greater than pretest scores on interpretation and production. Significant gains in linguistic development were thus evidenced not only at the time of the immediate posttests but also at the time of the delayed tests for both experimental groups on both interpretation and production tasks.

Our second research question was: If both instructional conditions lead to improved performance, is the performance evidenced by learners in PI and MOBI groups comparable? A one-way ANOVA performed on pretest scores revealed no significant differences between groups for either interpretation, $F(2, 42) = 0.31$, $p = .74$, or production, $F(2, 42) = 0.68$, $p = .51$. Therefore, any differences on posttests found between groups can be attributed to treatment

**Table 5.** Mean percentage correct for production tests

| | | Preproduction | | Postproduction | | Delayed production | |
|---|---|---|---|---|---|---|---|
| Group | *N* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| PI | 15 | 1.96 | 5.73 | 30.88 | 38.10 | 26.36 | 41.07 |
| MOBI | 14 | 0.56 | 2.15 | 69.93 | 40.32 | 31.55 | 37.20 |
| Control | 14 | 2.49 | 5.24 | 13.10 | 27.81 | 15.80 | 31.01 |

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

**Table 6.** ANOVA results for time: production tests

| Source of variability | *df* | Sums of squares | Mean squares | *F* | *p* | Partial eta$^2$ | Power |
|---|---|---|---|---|---|---|---|
| Pretest to immediate posttest | | | | | | | |
| PI | 1 | 6,688.95 | 6,688.95 | 8.80* | .01 | .37 | .79 |
| MOBI | 1 | 36,095.37 | 36,095.37 | 45.28* | .00 | .76 | 1.00 |
| Pretest to delayed test | | | | | | | |
| PI | 1 | 4,985.66 | 4,985.66 | 6.18* | .03 | .31 | .64 |
| MOBI | 1 | 6,706.35 | 6,706.35 | 9.46* | .01 | .42 | .81 |
| Immediate posttest to delayed test | | | | | | | |
| PI | 1 | 324.30 | 324.30 | 1.68 | .22 | .11 | .23 |
| MOBI | 1 | 9,190.20 | 9,190.20 | 11.43* | .00 | .47 | .88 |

*$p < .05$.

effects rather than to preexisting differences. Pretest, immediate posttest, and delayed test scores were submitted to a 3 × 3 repeated-measures ANOVA for interpretation. Group was the between-subjects factor and time was the within-subjects factor. Significant main effects were found for group and time and a significant interaction was found for group and time (see Table 7). A post hoc Scheffé revealed that the main effect for group was due to the following contrasts: PI better than control ($p < .001$) and MOBI better than control ($p < .01$). There was no significant difference between PI and MOBI ($p = .67$). In sum, analyses of the interpretation data indicate that both experimental conditions had a significant positive effect on linguistic development compared to the control group, but did not differ from each other, as seen in Figure 2. Thus, the performance evidenced by PI and MOBI groups appears to be generally comparable.

To obtain a clearer picture of the interaction between group and time, scores for the interpretation immediate posttest and delayed test were each submitted to a univariate ANOVA, with group as the between-subjects factor. For the interpretation immediate posttest, a significant main effect was found for group, $F(2, 42) = 29.23$, $p < .001$. A post hoc Scheffé revealed that this effect was due

**Table 7.** ANOVA results for between-subjects comparison for interpretation

| Source | *df* | Sums of squares | Mean squares | *F* | *p* | Partial eta$^2$ | Power |
|---|---|---|---|---|---|---|---|
| Group | 2 | 17,133.57 | 8,566.78 | 11.47* | .00 | .36 | .99 |
| Time | 2 | 49,531.86 | 24,765.93 | 64.63* | .00 | .62 | 1.00 |
| Group × Time | 4 | 18,032.34 | 4,508.08 | 11.76* | .00 | .37 | 1.00 |
| Residual | 80 | 30,657.57 | 383.22 | | | | |

*$p < .05$.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

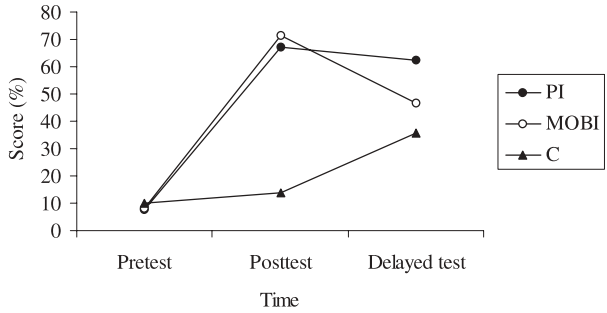



**Figure 2.** Interpretation scores.

to the fact that the scores of both PI and MOBI were significantly greater than those of the control ($p < .001$ and $p < .001$, respectively). The univariate ANOVA for the delayed interpretation test scores, however, showed no significant main effect for group, $F(2, 40) = 2.68$, $p = .08$. Therefore, although PI and MOBI performed better than control on the posttest, this difference between groups was not maintained on the delayed test.

Pretest, immediate posttest, and delayed production test scores were also submitted to a 3 × 3 repeated-measures ANOVA. Group was the between-subjects factor and time was the within-subjects factor. Significant main effects were found for group and time and a significant interaction was found (see Table 8). A post hoc Scheffé revealed that the main effect for group was due to the following contrast: MOBI better than control ($p < .05$). There was no significant difference between PI and control ($p = .55$) or between PI and MOBI ($p = .31$). A visual inspection of Figure 3 shows that MOBI made the greatest improvement, followed by PI and, finally, control. Interestingly, although there was no differential effect between PI and MOBI, only MOBI led to significantly greater linguistic development compared to control.

As with interpretation, further analyses were performed to clarify the interaction between group and time: Scores for the production immediate post-

**Table 8.** ANOVA results for between-subjects comparison for production tests

| Source | *df* | Sums of squares | Mean squares | *F* | *p* | Partial eta$^2$ | Power |
|---|---|---|---|---|---|---|---|
| Group | 2 | 11,069.47 | 5,534.74 | 3.47* | .04 | .15 | .62 |
| Time | 2 | 29,667.41 | 14,833.71 | 27.36* | .00 | .41 | 1.00 |
| Group × Time | 4 | 12,254.32 | 3,063.58 | 5.65* | .00 | .22 | .97 |
| Residual | 80 | 43,370.51 | 542.13 | | | | |

*$p < .05$.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

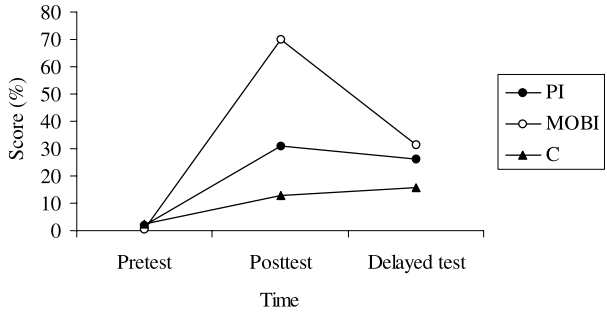



**Figure 3.** Production scores.

test were submitted to a univariate ANOVA. A significant main effect was found for group, $F(2, 42) = 9.53$, $p < .001$. A post hoc Scheffé revealed that this effect was due to the scores of MOBI being significantly greater than both PI ($p < .05$) and control ($p < .01$). A univariate ANOVA run on the delayed production test scores, however, showed no significant main effect for group on the delayed test, $F(2, 40) = 0.668$, $p = .52$. Therefore, although MOBI outperformed both PI and control on the production immediate posttest, this difference between groups was not maintained on the delayed test.

In sum, the results indicate the following regarding the second research question, related to the comparability of the performance of the experimental groups: On the interpretation task, both PI and MOBI outperformed control, whereas on the production task, only MOBI outperformed control. Moreover, neither PI nor MOBI outperformed the other on either task. Thus, overall, it appears that MOBI performs at least as well as PI.

## DISCUSSION

The present study investigated the effects of MOBI compared to PI. The study's first research question was: Do PI and MOBI lead to improved performance on interpretation and production of Spanish direct object pronouns? Our results indicate that for both experimental groups, immediate and delayed test scores were significantly higher than pretest scores for interpretation and production. Thus, the response to our first research question is affirmative: PI and MOBI both lead to improved performance for interpretation and production of Spanish direct object pronouns. The fact that MOBI evidenced significant losses from the immediate posttests to the delayed tests is a result that might appear contradictory to this conclusion. This might be interpreted in one of at least two ways. First, it is possible that participants in MOBI showed a regression to the mean, given their high scores on the immediate posttests. On the other hand, it might be that participants in PI are in fact evidencing more stable gains. Although further delayed tests would have elucidated this

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

issue, the observation that PI produces stable effects is consistent with previous PI research (Cadierno, 1995; Farley, 2001a, 2000b; VanPatten & Cadierno, 1993a, 1993b). Overall, despite MOBI's loss from posttests to delayed tests, MOBI—like PI—evidenced significant gains from pretests to delayed tests, and, therefore, both instructional conditions led to improved performance.

Our second question was: If both instructional conditions lead to improved performance, is the performance evidenced by learners in PI and MOBI groups comparable? Repeated-measures ANOVAs showed that, for interpretation, PI and MOBI outperformed control and that neither PI nor MOBI outperformed the other. For production, only MOBI outperformed control and, again, there was no difference between PI and MOBI. However, a closer look at performance on each posttest with one-way ANOVAs conducted on the immediate posttest and delayed test scores reveals a more complex picture. For the interpretation immediate posttest, both experimental groups outperformed the control group but did not differ significantly from one another. For the production immediate posttest, on the other hand, MOBI outperformed both control and PI, whereas PI did not outperform control. For the delayed tests—for both interpretation and production—no differences were found between any groups. Overall, the results that address our second research question reveal that MOBI and PI performed similarly on interpretation measures, and only MOBI outperformed control on production.

It is important to consider these findings in relation to other studies that have examined the effects of PI and MOBI (e.g., Farley [2001a, 2001b] and Benati [2005]). Given that analyses for those studies were based on overall repeated-measures ANOVAs, we will consider only these analyses of our data in comparing results. On the one hand, results related to our first research question are consistent with both Farley's and Benati's findings; in all four studies, both PI and MOBI have a significant effect on linguistic performance. On the other hand, results related to our second research question are partially inconsistent with those found in their studies. Our results are similar to Farley's (2001b) study in that neither of our experimental groups outperformed the other on either interpretation or production, but they are different from Farley (2001a), given that his PI group outperformed his output-based group on interpretation. Due to the fact that delayed tests were not administered in Benati, the analyses reported from the current study are not directly comparable. Thus, in order to make a direct comparison, we carried out $3 \times 2$ repeated-measures ANOVAs on the pretests and immediate posttests only. Whereas Benati's PI group outperformed his MOI group on interpretation and both groups were statistically equivalent on production, our MOBI group—under this analysis—outperformed PI on production, and both groups were statistically equivalent on interpretation.[10] Although there are several factors that introduce differences among all four studies (e.g., age of participants, target structure, first language–L2 pairings, amount of practice, classroom vs. computer-administered instruction), there is no apparent underlying theoretical explanation based on these factors that might account for these differences. Nevertheless, the four studies' designs are similar and the results are fairly

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

consistent: The common finding for all these studies is that both PI and MOI/MOBI groups made significant gains for both interpretation and production.

What theoretical approach best explains both our results and those of previous studies? One consistent finding has been that PI leads to improved linguistic performance. The mechanisms underlying this linguistic development are readily explained by VanPatten's model of input processing and have been discussed in previous related research (DeKeyser et al., 2002; VanPatten, 1996, 2002, 2004a, 2004b; VanPatten & Cadierno, 1993a, 1993b). What merits further discussion is what mechanisms best account for the improved performance of MOBI, including greater benefits on the immediate production posttest as compared to PI. Our discussion of these mechanisms will be based on two perspectives: (a) one in which input directly leads to acquisition but output might only affect acquisition indirectly or lead to skill-specific knowledge and (b) one in which both input and output can directly affect acquisition.

Under the position that input—but not output—directly affects acquisition, a conservative stance might be that input alone leads to linguistic development. This is consistent with Farley's (2001b) interpretation of his findings, in which equivalent benefits obtained for his PI and MOI groups. He attributed the linguistic improvement of the MOI group in his study to the incidental input that participants received in the treatment. Farley argued that because the target form was in the initial position in output produced by students and teachers, this might have served as structured input, "an element similar to PI—input in which the target form is more salient and more likely to be processed" (p. 84). Participants in our study received similar input through feedback, which also might have been effectively equivalent to structured input. This structured input alone, however, cannot account for our results. Because they received this input only for incorrect answers, our MOBI participants received fewer instances of structured input compared to the PI group; thus, if input were the only causative factor, we would expect the PI group to outperform the MOBI group. Our results, however, indicate that PI never outperformed MOBI.

Still under this same broad perspective, another alternative is that only input directly affects acquisition, but output might lead to skill-specific knowledge or might indirectly promote linguistic development. VanPatten (2002, 2004b) held the position that only input directly affects acquisition, and VanPatten and Cadierno (1993a) stated that the best explanation for their results was that their output-based TI resulted in a "different knowledge system" (p. 238), which was reflected on the production tests. Thus, the interpretation of the results of the present study might be that both experimental groups evidenced linguistic gains on interpretation and production tasks due to the input in their respective treatments, whereas the MOBI group had superior performance on the production tasks due to a task effect from the output practice that only they received. This explanation, however, does not account for why the MOBI group performed as well as the PI group on interpretation tasks even though they had less input. Assuming that only input directly affects acquisition but that output might indirectly affect linguistic development, it

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

might be that—at least under certain conditions—output encourages more noticing of the target form in the input and therefore makes it more likely to be processed. There are at least two ways in which this might occur. First, the juxtaposition of one's own output with additional input might lead to noticing of this input. In fact, VanPatten (2004b) suggested that in negotiation, "the immediacy of juxtaposing one's output with another's input may trigger noticing that is useful for making form-meaning connections" (p. 13). In our MOBI treatment, the juxtaposition of learner output with subsequent input provided in the feedback might have triggered noticing and thereby led to the linguistic gains evidenced by the output-based group. Second, as suggested by an anonymous *SSLA* reviewer, the fact that learners' own output in this treatment condition triggered the individualized feedback might have made this additional input psycholinguistically more salient for the learners and, thus, more likely to be noticed. These latter explanations are consistent with our findings and, therefore, plausible.

The second broad perspective from which these results might be understood is one in which output itself—as well as input—might lead directly to acquisition. Two of the functions posited by Swain's output hypothesis (1993, 1995, 1998) seem to allow for a direct effect of output on the developing system: hypothesis testing and noticing. Swain (1993) and de Bot (1996) contended that producing language encourages hypothesis testing. In Gass' (1997) model of SLA, hypothesis testing "constitutes a means for output . . . to serve as feedback into the intake component" (p. 7) and, thus, it appears that output might be able to affect acquisition directly (i.e., without dependence on further input). Regarding noticing, it has been proposed that through output, learners might notice a gap between what they want to communicate and what they are able to produce. Swain (1998) suggested that "[t]his may trigger cognitive processes that might *generate* [italics added] linguistic knowledge that is new for the learner or *consolidate* [italics added] the learner's existing knowledge" (p. 67). Ellis (2005) discussed processes that may occur during L2 production, including processes that are consistent with the functions of the output hypothesis described previously. He maintained that these processes, among others, may affect implicit learning during output. It is not unreasonable to hold that mechanisms such as hypothesis testing and noticing a gap may lead to changes in the developing system through the generation or consolidation of knowledge.

The underlying mechanisms of such generation or consolidation of knowledge might be understood as the establishment or strengthening of form-meaning connections. The latter idea is consistent with de Bot (1996), in which he provided an initial psycholinguistic account of the functions of the output hypothesis position. He held that connections might be strengthened through output when form-meaning connections are made and "the system does not get error messages as a result of this connection" (p. 549). On the other hand, "if what is produced and what is correct do not match according to the internal norm, negative feedback will hamper the development of the connection" (p. 549). In fact, the establishment and strengthening of form-meaning connec-

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

tions might be a common element that underlies the effectiveness of certain types of input-based and output-based practice. Both VanPatten (1996, 2004b) and Swain (1993, 1995) asserted that when processing input, learners are primarily focused on extracting meaning, which might lead to a lack of processing of forms. Similarly, both have suggested a manner in which learners might be led to process forms. PI pushes learners "away from the less-than-optimal strategies" and requires them "to rely on form or sentence structure to interpret meaning" (VanPatten, 2004b, p. 37), thereby moving from lexical to grammatical processing. Swain (1995) claimed that producing output might "stimulate learners to move from the semantic, open-ended, non-deterministic, strategic processing prevalent in comprehension to the complete grammatical processing needed for accurate production" (p. 128). Thus, both perspectives advocate language practice that encourages learners to process forms correctly for meaning and thereby make correct form-meaning connections. Given results of the current study along with results of the research reviewed, it is likely that both input-based and output-based activities can do precisely this.

If one were to interpret the current study's findings from this perspective, in which both input and output can directly affect acquisition through the establishment of form-meaning connections, the analysis of our results is straightforward: In PI, input-based practice leads to form-meaning connections evidenced in both interpretation and production test gains, whereas in MOBI, learners might have been able to establish or strengthen form-meaning connections via both output-based practice and the input they received in instruction and feedback. This account would explain why the MOBI group performed as well as or better than the PI group on all tests. Although this account is consistent with our results and is theoretically plausible, because our output-based treatment did not completely isolate the variable of output (i.e., in the absence of input in feedback or instruction), we cannot positively conclude that the element of output practice in and of itself affected the developing system. What we can conclude is that, in our study, an output-based instruction that included some input led to at least as much linguistic development as an input-based instruction that did not include output. This conclusion is consistent with some (Cheng, 2002; Collentine, 1998; Farley, 2001b; Izumi & Bigelow, 2000), but not all (Benati, 2005; Farley, 2001a) other research in which the tasks—irrespective of mode—seemed to encourage a focus on both form and meaning. These studies' results, like ours, indicated that the output-based groups made similar, if not greater, gains in linguistic development compared to the input-based groups, across a variety of tasks.

Considering the results of the current study along with those of PI studies, output studies, and studies examining the relative effects of input and output, what larger conclusions can be drawn that are significant to SLA theory? This study contributes to a growing body of literature that suggests that output—as well as input—can promote linguistic development and might constitute a direct path to acquisition via the establishment and strengthening of form-meaning connections. The current study is particularly important to this body

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

of literature, given the various methodological improvements included as part of the research design.

It is important to point out that we believe that neither the current study nor the other studies we have discussed is able to definitively shed light on whether the linguistic gains evidenced are attributable to effects on acquisition, or on a different system of learned linguistic knowledge that is available only for monitoring, or both. This is because none of the studies employed a methodological design that elucidated learners' psycholinguistic processing, given that what was measured in these studies was linguistic development on fairly controlled performance assessments. In fact, DeKeyser et al. (2002) claimed that "very little if any research on 'PI' can even claim to address acquisition and not just the learning of monitored knowledge" (p. 819). If progress is to be made with regard to dissociating these posited systems, then what is needed in SLA is a valid, agreed-upon method for accomplishing this. The use of reaction time measures and neurolinguistic assessments might provide data that are more likely to elucidate this larger theoretical issue.

Several other considerations should be explored in relation to the present study. First, it is curious that although both PI and MOBI groups evidenced linguistic development on delayed tests relative to pretests, between-group analyses revealed that the experimental groups were not significantly different from each other, or from the control group, on delayed tests. The fact that MOBI was no longer superior to PI for production at the time of the delayed test appears to be due to the loss of MOBI between immediate and delayed production tests. The lack of a significant difference among all three groups on the delayed tests is likely attributable to the fact that the control group evidenced gains on the interpretation and production delayed tests, although these gains were significant only for interpretation. If the control group had not shown these gains, it is likely that the experimental groups would have outperformed the control group on this measure.

This leads to a second consideration: What accounts for the gains made by the control group? A significant factor in the current study's research design is that the control group was exposed to the target forms in meaningful contexts in reading passages. This was not the case in the vast majority of previous research, which overwhelmingly used nonexposure control groups. Although such studies might conclude that participants in their treatments made gains compared to participants who were not exposed to the target form, they cannot say whether these treatments would have led to greater gains compared to exposure to target forms in a meaningful context or even simple exposure to the form. For the current study, it is likely that exposure to the target form, in combination with test effects, or prior knowledge, or both, led to the control group's gains on delayed tests. After exposure to the target form in a meaningful context, participants might have had the opportunity to make form-meaning connections on the immediate posttests, which led to their gains on the delayed interpretation test.

Prior knowledge is another issue that should be examined more carefully. Although the current study employed a pretest cutoff that was more stringent

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

than those of most previous studies and although first-semester Spanish students were selected as participants, it was, nevertheless, impossible to completely eliminate the potential influence of prior knowledge of the target form in Spanish or of similar forms in other Romance languages due to participants' prior language study.[11] As previously reported, there were no differences among groups with regard to prior Romance language experience. Nevertheless, it might have affected all treatment groups to some degree and would have been most visible in the gains of the control group. It is probable, however, that prior knowledge played less of a role in our study than in other studies. The fact that participants in other studies apparently had more prior knowledge might partially explain the differences in results. Specifically, in other studies, PI maintained its advantage over control even on delayed tests. For example, this was the case in VanPatten and Cadierno (1993a, 1993b), in which participants were second-year Spanish students and, thus, had likely studied the target form previously at the university level and were, in a sense, recycling previously studied concepts. Thus, their gains might have been maintained in part due to the interaction between previous knowledge and the treatments. Our participants, on the other hand, had not yet formally studied the target form at the university; thus, their lower prior knowledge in combination with the treatments might have been sufficient to bring about advantages over control on the immediate posttests; however, these effects were not as strong as effects in other studies, in which advantages were maintained on delayed tests.

Finally, we should consider the limitations of this study. First, as stated previously, our MOBI group was not a pure output group because these participants were provided input in the instruction as well as in the feedback. As a consequence, our study is unable to reveal the effects of output alone. Similarly, in Farley (2001a, 2001b), participants in the MOI group were provided with input when they heard the correct answers given by their classmates. It appears that this was the case in Benati (2005) as well. One advantage of a computer-based treatment is that input from other learners can be eliminated and that feedback can be individualized. Thus, in the future, it would be interesting to create a MOBI that is administered using a computer and that indicates only whether learners' answers are correct or incorrect. Indeed, if the goal is to isolate the role of output in SLA, such a treatment should be provided without input from grammatical explanation. Such a treatment might have implications regarding the theoretical role of output. At the same time, this sort of condition might not be perceived as instructionally valid by learners. A second limitation of the current study is that we were unable to track participants' responses and the feedback they received. Thus, we were unable to monitor participants' progress throughout the treatment or to determine the exact amount of input provided through feedback to individuals in the MOBI group. Future research should be able to improve on both of these aspects of study design. It is unlikely, however, that technological advances will allow for evaluation of oral output in the near future. Thus, the creation

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

of a pure output group that incorporates individualized feedback to oral output in a computer-administered paradigm will remain problematic. A third limitation to our study is that although attempts were made to ensure sufficient vocabulary knowledge for completion of the study, we cannot be sure that our beginning-level participants retained new vocabulary that was practiced approximately 2 weeks prior to treatment. Finally, the delayed tests were administered only 1 week after treatment. We decided not to administer additional delayed tests in this study because the participants' Spanish classes were to begin studying direct object pronouns only a few days after the 1-week delayed test; therefore it would have been impossible to control for this intervening variable. It would have been interesting, however, to have administered further posttests in order to see whether any differences between groups would have been observed at this later point in time.

## CONCLUSION

As pointed out by many researchers (e.g., Gass, 1997; VanPatten, 2002), it is clear that input is essential to SLA. Nevertheless, in this study, the MOBI group performed at least as well as the PI group, as evidenced by the findings that (a) both groups made significant gains on immediate and delayed posttests as compared to pretests, for both interpretation and production, and (b) only the MOBI group outperformed the control group on production overall. For the MOBI group, however, there was a significant loss between the immediate and delayed posttests for both interpretation and production, which led to no differences between groups at the time of the delayed tests. We thus conclude that MOBI can, like PI, lead to linguistic development, at least when practice is meaningful and leads learners to make form-meaning connections. Pedagogically, the results of our study seem to support the use of meaningful output practice as well as meaningful input-based practice in the L2 classroom environment as a means for building fluency and accuracy, not only in later stages of instruction but also during early stages of instruction of new forms. Future research should be carried out to replicate these findings and to determine what other types of input- and output-based methods are effective for SLA. Likely candidates are those methods that provide communicative and meaningful output-based practice, with a focus on form as well, such as in task-based instruction (Bygate, Skehan, & Swain, 2001; Nunan, 1989). Additionally, future research should be designed to completely isolate the effects of output practice and thus shed light on its theoretical status in SLA.

(*Received 12 July 2005*)

### NOTES

1. Following VanPatten (2004a), we use the term *acquisition* to mean "the development of some underlying competence on which skills in language use depend. . . . This competence is also known

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

as an underlying mental representation, the developing system, the interlanguage, and other terms used in the SLA literature" (p. 29).

2. VanPatten and colleagues have used interpretation rather than comprehension tasks. For them, and for the current study, interpretation implies processing of the target form such that its critical information is understood. In the case of direct object pronouns, this means that "subjects would have to process noun phrases as objects or subjects" (Sanz & Van Patten, 1998, p. 271). Some other studies (see Table 2) use comprehension tasks, some of which focus on morphological aspects of the form (i.e., number or gender) but do not require interpretation of the form as defined by Sanz and VanPatten. In some cases, these tasks can be completed simply by matching the number or gender of the form to an English translation, without understanding the form's meaning. When discussing each study, we will follow their terminology, but the reader should consider that interpretation and comprehension tasks do not necessarily measure the same competence.

3. VanPatten (2002), who based his argumentation on Cheng's (1995) dissertation, interpreted the results from the *estar*-only data as supporting the effectiveness of PI, but did not cite Cheng (2002). DeKeyser et al. (2002) noted that "Cheng (2002) appears to supersede the overly strong claim about the benefits of PI made by Cheng (1995)" (p. 818).

4. In an error-correction exercise provided in Erlam (2003), the basis for making a grammaticality judgment appears to be the placement of the direct object pronoun. Thus, the learners were not required to interpret the target form meaningfully in order to answer correctly.

5. Participants had an average of 0.8 years (range: 0–4 years) of Spanish instruction in high school and an average of 0.3 years (range: 0–2 years) prior to high school. (This is with the exception of one participant, who indicated that she had Spanish once a week in elementary school for 6 years.) Fourteen participants indicated previous study of French (the only other modern Romance language reported), with an average of 2.4 years of study. Previous PI studies either did not report prior language experience of participants or had seemingly arbitrary cutoff criteria based on this experience. Although our original intent was to exclude participants with any previous exposure to any Romance language in order to rule out any prior exposure to the target form or analogous forms in other Romance languages, we found that a no-exposure criterion was so strict that it would have reduced the total subject pool to 12.

6. Analyses were also run using the 80% cutoff criterion employed by VanPatten and Cadierno (1993a, 1993b), and the same pattern of significant differences was found as with 33%.

7. These materials are included in their entirety in Cadierno's (1993) dissertation.

8. To investigate learner-internal variables, a working memory test was administered immediately before the treatment and think-aloud protocols were collected during the treatment. Preliminary results of the think-aloud data have been presented by the authors (Bowden & Morgan-Short, 2004) and will be discussed in a manuscript currently in progress.

9. Test versions varied from 9 to 12 critical items due to the fact that some items, which contained errors, were eliminated from analysis. Percentage scores were used so that scores across test versions could be directly compared.

10. The results of these repeated-measures ANOVAs are as follows. For interpretation, significant main effects were found for group, $F(2, 42) = 23.75$, $p < .001$, and time, $F(1, 42) = 137.40$, $p < .001$, and a significant interaction was found for group and time, $F(2,42) = 27.63$, $p < .001$. A post hoc Scheffé revealed that the main effect for group was due to the following contrasts: PI better than control, $p < .001$, and MOBI better than control, $p < .001$. There was no significant difference between PI and MOBI, $p = .88$. For production, significant main effects were found for group, $F(2, 42) = 8.38$, $p < .01$, and time, $F(1, 42) = 46.75$, $p < .001$, and a significant interaction was found for group and time, $F(2,42) = 10.48$, $p < .001$. A post hoc Scheffé revealed that the main effect for group was due to the following contrasts: MOBI better than control, $p < .01$, and MOBI better than PI, $p < .05$. There was no significant difference between PI and control, $p = .45$.

11. Although not part of the original research questions, given that prior language study could not be completely controlled and possibly played a role in our results, post hoc analyses were carried out in order to investigate any influence of this variable on observed linguistic development. Correlations were run between linguistic development and years of Romance language study. A positive correlation was found with gains on the posttests and delayed production tests, $r = .33$, $p < .05$ and $r = .35$, $p < .05$, respectively. Thus, prior Romance language study appears to be positively related to improved performance on production. We can only speculate as to why Romance study is not correlated with interpretation gains. Future studies should collect and report information on previous language study in order to investigate potential effects on linguistic development.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

## REFERENCES


Allen, L. Q. (2000). Form-meaning connections and the French causative: An experiment in processing instruction. *Studies in Second Language Acquisition*, *22*, 69–84.

Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research*, *5*, 95–127.

Benati, A. (2005). The effects of processing instruction, traditional instruction, and meaningful-based output instruction on the acquisition of English past simple tense. *Language Teaching Research*, *9*, 67–93.

Bowden, H. W., & Morgan-Short, K. (2004, May). *Attention, instructional conditions, and second language development*. Paper presented at the meeting of the American Association of Applied Linguistics, Portland, OR.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. London: Longman.

Cadierno, T. (1993). Explicit instruction in grammar: A comparison of input based and output based instruction in second language acquisition. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1990). *Dissertation Abstracts International*, *53*, 3889A–3890A.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *Modern Language Journal*, *19*, 179–193.

Cheng, A. C. (1995). *Grammar instruction and input processing: The acquisition of Spanish ser and estar*. (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1995). *Dissertation Abstracts Internationsl*, *56*, 3485A–3486A.

Cheng, A. C. (2002). The effects of processing instruction on the acquisition of *ser* and *estar*. *Hispania*, *85*, 308–323.

Collentine, J. (1998). Processing instruction and the subjunctive. *Hispania*, *81*, 576–587.

Collentine, J. (2002). On the acquisition of the subjunctive and authentic processing instruction: A response to Farley. *Hispania*, *85*, 879–888.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, *46*, 529–555.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*, 195–221.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). New York: Cambridge University Press.

DeKeyser, R. M., Salaberry, M. R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "Processing instruction: An update." *Language Learning*, *52*, 805–823.

DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practices. *Language Learning*, *46*, 613–642.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.

Erlam, R. (2003). Evaluating the relative effectiveness of structured-input and output-based instruction in foreign language learning. *Studies in Second Language Acquisition*, *25*, 559–582.

Farley, A. P. (2001a). Authentic processing instruction and the Spanish subjunctive. *Hispania*, *84*, 289–299.

Farley, A. P. (2001b). Processing instruction and meaning-based output instruction: A comparative study. *Studies in Applied Linguistics*, *5*, 57–93.

Farley, A. P. (2002). Processing instruction, communicative value, and ecological validity: A response to Collentine's defense. *Hispania*, *85*, 889–895.

Farley, A. P. (2004). The relative effects of processing instruction and meaning-based output instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 143–168). Mahwah, NJ: Erlbaum.

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

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*, 541–577.

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. *Studies in Second Language Acquisition*, *21*, 421–452.

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

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *74*, 440–464.


https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

Nagata, N. (1998). Input vs. output practice in educational software for second language acquisition. *Language Learning & Technology, 1*, 23–40.
Nunan, D. (1989). *Designing tasks for the communication classroom*. New York: Cambridge University Press.
Salaberry, M. R. (1997). The role of input and output practice in second language acquisition. *Canadian Modern Language Journal*, *53*, 422–451.
Sanz, C., & VanPatten, B. (1998). On input processing, processing instruction, and the nature of replication tasks: A response to M. Rafael Salaberry. *Canadian Modern Language Journal*, *54*, 263–273.
Schwartz, B. (1993). On explicit and negative data effecting and affecting *competence* and *linguistic behavior*. *Studies in Second Language Acquisition*, *15*, 147–163.
Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, *50*, 158–164.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of H. 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.
VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Westport, CT: Ablex.
VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*, 755–803.
VanPatten, B. (2004a). Input and output in establishing form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 29–47). Mahwah, NJ: Erlbaum.
VanPatten, B. (2004b). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–32). Mahwah, NJ: Erlbaum.
VanPatten, B., & Cadierno, T. (1993a). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–243.
VanPatten, B., & Cadierno, T. (1993b). Input processing and second language acquisition: A role for instruction. *Modern Language Journal*, *77*, 45–57.
VanPatten, B., Lee, J., & Ballman, T. (2000). *¿Sabías que. . .? Beginning Spanish* (3rd ed.). Boston: McGraw-Hill.
VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, *18*, 495–510.
VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. Eckman, D. Highland, P. Lee, J. Mileham, & R. Rutkowski (Eds.), *Second language acquisition: Theory and pedagogy* (pp. 169–185). Mahwah, NJ: Erlbaum.
VanPatten, B., & Wong, W. (2004). Processing instruction and the French causative: Another replication. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 97–118). Mahwah, NJ: Erlbaum.

# APPENDIX A

## SAMPLE PI PRACTICE ACTIVITIES

1. Me llama frecuentemente mi hermana. ("My sister calls me frequently.")

   Who calls whom? a. I call my sister. b. My sister calls me.

2. Read the statement that follows and select the ones that you think are typical.

   Los parientes . . . ☐ nos visitan. ☐ nos aman. ☐ nos atacan.

   ("Relatives . . . ☐ visit us. ☐ love us. ☐ attack us.")

*Note*. Translations appear in parentheses but were not on original task. After completion, participants who answer correctly see: "Correct"; those who answer incorrectly see: "Sorry, that's incorrect."

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

# APPENDIX B

## SAMPLE MOBI PRACTICE ACTIVITIES

1. Think about the difference between the following two sentences in English.

   a. I call my sister. b. My sister calls me.

   Sentence (a) in Spanish is "Llamo a mi hermana."

   How do you think you might say sentence (b)? _____ llama mi hermana.

*Note*. After completion, participants who answer correctly see: "Correct;" those who answer incorrectly see: "Sorry, that's incorrect. The correct answer is 'me'."

2. Complete the sentences below indicating how relatives do or do not act toward us.

   a. Los parientes _____ (visitar).

   ("Relatives _____ (to visit).")

   b. Los parientes _____ (amar).

   ("Relatives _____ (to love).")

   c. Los parientes _____ (atacar).

   ("Relatives _____ (to attack).")

*Note*. Translations appear in parentheses but were not on original task. After completion, participants who answer correctly see: "Correct;" those who answer incorrectly see: "Sorry. The correct answer is 'nos visitan' ('nos aman', 'no nos atacan')."

# APPENDIX C

## SAMPLE CONTROL ACTIVITY—EXCERPT FROM "UN TALENTO ESPECIAL"

Los abuelos maternos de Sara son mexicanos y los quiere mucho. Viven en San José y cuando Sara viaja a California, siempre los visita.

Su abuela se llama Concepción y es una persona muy especial. Tiene una habilidad psíquica (puede ver eventos del futuro y del pasado) pero no la usa con mucha frecuencia. Dice que es un regalo de Dios y debe usarla con mucho cuidado.

1. ¿Dónde viven los abuelos de Sara?
2. ¿Cuándo los visita?
3. ¿Cómo se llama la abuela de Sara?
4. ¿Qué tipo de persona es la abuela?
5. ¿Qué tipo de habilidad tiene la abuela de Sara?
6. ¿Qué dice la abuela de su habilidad?

*Note*. No feedback was provided.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

# APPENDIX D

## SAMPLE INTERPRETATION TEST ITEM

Task explanation: Participants hear a sentence in the L2 and see three pictures. Participants indicate which picture corresponds to the sentence heard.

Experimental sentence: Lo carga María.
("María carries him.")



A



B



C

Answer choices: A B C Don't know.

*Note*. Translation of the experimental sentence appears in parentheses but was not heard by participants.

# APPENDIX E

## SAMPLE PRODUCTION TEST ITEM

Task explanation: Participants see two pictures and complete a written sentence in the L2 corresponding to the actions depicted in the pictures.





https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press

Experimental sentence to be completed:

1. Pancho besa a Tita y después (abrazar) _____.
   ("Pancho kisses Tita and then (to hug) _____.")

*Note*. Translation of the experimental sentence appears in parentheses but was not provided for participants.

https://doi.org/10.1017/S0272263106060025 Published online by Cambridge University Press