SSLA, **27**, 361–386. Printed in the United States of America.
DOI: 10.1017/S0272263105050163

INCIDENTAL FOCUS ON FORM AND SECOND LANGUAGE LEARNING

Shawn Loewen
University of Auckland

Incidental focus on form overtly draws learners' attention to linguistic items as they arise spontaneously—without prior planning—in meaning-focused interaction. This study examined the effectiveness of incidental focus on form in promoting second language (L2) learning. Seventeen hours of naturally occurring, meaning-focused L2 lessons were observed in 12 different classes of young adults in a private language school in Auckland, New Zealand. A total of 491 focus-on-form episodes (FFEs) were identified and used as a basis for individualized test items in which participants who participated in specific FFEs were asked to recall the linguistic information provided in them. The results revealed that learners were able to recall the targeted linguistic information correctly or partially correctly nearly 60% of the time 1 day after the FFE, and 50% of the time 2 weeks later. Furthermore, successful uptake in a FFE was found to be a significant predictor of correct test scores. These results suggest that incidental focus on form might be beneficial to learners, particularly if they incorporate the targeted linguistic items into their own production.

Incidental focus on form is considered to be a potential way to achieve an integration of meaning-focused and form-focused activities in the second language (L2) classroom (Ellis, 2001). Recent research (Ellis, Basturkmen, &

This research was supported, in part, by a scholarship from the Foundation for Science, Research, and Technology, New Zealand. The author would like to thank the teachers and participants who kindly gave of their time to participate in this study. The author would also like to thank Rod Ellis, Catherine Elder, Helen Basturkmen, Rosemary Erlam, Jenefer Philp, and the anonymous *SSLA* reviewers for their valuable input and feedback on this study. Earlier versions of this paper were presented at the annual meeting of the American Association of Applied Linguistics in 2002 (Salt Lake City, UT) and at AILA 2002 (Singapore).

Address correspondence to: Shawn Loewen, Department of Applied Language Studies and Linguistics, University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: s.loewen@auckland.ac.nz.

© 2005 Cambridge University Press 0272-2631/05 $12.00

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Loewen, 2001a, 2001b; Lyster, 1998a, 1998b; Lyster & Ranta, 1997) has investigated the occurrence of incidental focus on form; however, these studies have been primarily descriptive in nature and have investigated the effectiveness of incidental focus on form only in relation to learner uptake (i.e., learners' utterances in response to the feedback provided to them). These studies suggest that uptake might be a possible indication of learning, although they admit that uptake cannot necessarily be taken as evidence of learning. The present study extends previous research by examining learners' ability to subsequently recall the linguistic feedback provided in incidental focus on form on individualized tests.

INCIDENTAL FOCUS ON FORM

Whereas meaning-focused instruction involves tasks and activities in which participants are primarily concerned with message exchange (Richards & Rodgers, 2001; Savignon, 2000), form-focused instruction encompasses "any pedagogical effort to draw learners' attention to language either implicitly or explicitly" (Spada, 1997, p. 73). Form-focused instruction has been divided into a focus on formS and a focus on form (Long, 1991, 1996). Focus on formS is defined as instruction in which syllabi and lessons are based on linguistic items, with the primary goal being to teach those items (Long, 1991). A focus on form differs in that it "overtly draws students' attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication" (Long, 1991, pp. 45–46). As a result, learners' attention is drawn to linguistic items only as the need arises and not in predetermined ways (Spada).

Although Long's (1991) original definition of focus on form stated that the attention arose incidentally, subsequent studies expanded the definition to include attention to form that was preplanned. Consequently, Ellis (2001) distinguished between planned and incidental focus on form. Planned focus on form involves targeting preselected linguistic items during a meaning-focused activity, either through input (e.g., input flood or input enhancement) or output (e.g., corrective feedback on errors in the use of pretargeted forms). In contrast, the linguistic items addressed in incidental focus on form arise spontaneously in the course of meaning-focused activities. Although both types of focus on form might be beneficial for learners (Doughty & Williams, 1998), their impact might vary. Planned focus on form has the advantage of providing intensive coverage of one specific linguistic item, whereas incidental focus on form provides extensive coverage, targeting many different linguistic items (Ellis et al., 2001a). Although the effectiveness of planned focus on form has been investigated in various contexts (e.g., Doughty & Williams; Long, Inagaki, & Ortega, 1998), fewer studies have examined the effects of incidental focus on form (Williams, 2001).

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

An example of incidental focus on form is shown in (1), taken from Ellis et al. (2001a). The focus-on-form episode (FFE), which includes all of the discourse that relates to one particular linguistic item, occurred during a role-play activity in which participants provided alibis for a fictitious crime. In the episode, the participant states where he was when the crime took place, and the teacher responds with a confirmation check that includes a correction of the missing definite article (i.e., a recast). The participant acknowledges the teacher's utterance and continues with his alibi while the teacher proceeds to elicit more information about the alibi. Thus, this FFE demonstrates a brief, spontaneous focus on a linguistic item within the context of a meaning-focused task.

(1) L: *I was in pub.*
T: *in the pub*?
L: *yeah* and I was drinking beer with my friend
T: which pub did you go to?

Potential Benefits of Incidental Focus on Form

In light of evidence that meaning-focused instruction might not result in learners reaching nativelike levels of accuracy (Genesee, 1987; Swain, 1985), applied linguists have called for an integration of meaning-focused and form-focused instruction (Ellis, 2001; Hulstijn, 1995; Skehan, 1998). Although not the only possibility, one way for such an integration to occur is through incidental focus on form. Incidental focus on form has appeal because it occurs within the context of meaning-focused activities, when the learners themselves experience difficulty with a linguistic item (Doughty & Williams, 1998; Long & Robinson, 1998). Incidental focus on form may cause learners to attend to form, meaning, and use in a single cognitive event, and this joint processing that takes place is proposed to facilitate "the cognitive mapping among forms, meaning, and use that is fundamental to language learning" (Doughty, 2001, p. 211).

Additionally, incidental focus on form can provide opportunities for learners to alternate their attention between processing language for meaning and for form. VanPatten (1990) argued that L2 learners might have difficulty attending simultaneously to meaning and form and that learners might opt for whichever pays the most dividends. With meaning-focused instruction, that would generally be meaning because it is central to the task. Incidental focus on form can provide a brief time-out from focusing on meaning and assist learners in noticing linguistic items in the input that might otherwise go unnoticed in entirely meaning-focused lessons (Ellis et al., 2001a; Schmidt, 2001; Skehan, 1998).

In addition to its ability to help induce noticing of linguistic items in the input—which is argued to be essential for L2 development (Gass, 1997; Schmidt, 1990, 1995, 2001)—incidental focus on form can also provide opportunities for learners to produce *pushed output* (Swain, 1985, 1995, 2000; Swain & Lapkin, 1995). Swain's output hypothesis claims that producing language serves L2

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

development in several ways, such as by promoting proceduralization (Lyster, 1998b) or fluency (de Bot, 1996; Skehan, 1998; Swain, 1995) in language production. As learners produce the L2, they become more proficient in the use of those linguistic forms. Another possible function of output is hypothesis testing (Gass; Skehan; Swain, 1995). As learners become aware of new linguistic forms, they might try them out in their own production, which can provide learners with feedback as to the correctness of their linguistic hypotheses. Additionally, the output itself and the potential feedback following it can serve as input that feeds back into the cognitive system and targets the specific areas being hypothesized by the learner (de Bot; Gass).

Finally, producing pushed output can force learners to process language syntactically rather than semantically (Gass, 1997; Skehan, 1998; Swain, 1995). If learners attend only to the meaning of the L2, they might not process the language as deeply, given that they do not necessarily need to understand every word and morpheme to get the gist of an utterance. Learners also sometimes indicate understanding when in fact they have not comprehended an utterance (Hawkins, 1985, cited in Swain, 1995). Swain claimed that pushed output forces learners to deeper, grammatical processing that potentially has a "significant role in the development of syntax and morphology" (p. 128). Similarly, de Bot (1996) argued that producing the linguistic form and making the cognitive connections for oneself is better than merely perceiving the forms.

Descriptive Studies

As was mentioned, several studies have investigated the occurrence of incidental focus on form in various contexts. For example, in a study of negative feedback in French immersion classes in Canada, Lyster (1998a, 1998b; Lyster & Ranta, 1997) found that 62% of participant errors were followed by some kind of teacher feedback. Ellis et al. (2001a, 2001b), in their study of meaning-focused lessons in a private language school in New Zealand, found that incidental focus on form occurred at the rate of one episode every 1.6 minutes. Finally, Loewen (2003), in an investigation of 12 classes in another private language school in New Zealand, found a range of 0.24 to 1.24 FFEs per minute. These studies suggest that incidental focus on form does occur in meaning-focused L2 classroom interaction, although the rate might be variable.

These descriptive studies have identified several characteristics of incidental focus on form, with an important distinction made regarding the instigation of attention to form. Much planned focus-on-form literature has investigated reactive focus on form, which occurs in response to learner errors; however, Ellis et al. (2001b) also identified preemptive focus on form, which occurs when either the teacher or a learner initiates attention to form, generally by raising a question, "even though no actual problem in production has arisen" (p. 414). Although both reactive and preemptive focus on form might be beneficial, learner topicalization of linguistic items in student-initiated focus

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

on form might be particularly useful, because learners are in a position to recognize and raise linguistic items that are problematic for them (Ellis et al.; Slimani, 1989). For example, Slimani's study found that learners reported a higher level of learning for linguistic items that they had initiated. Similarly, Ellis et al. found higher levels of uptake in student-initiated episodes than in teacher-initiated ones.

Once a focus on form has been instigated—either by a learner error or query—there is a response, often by the teacher. Broadly, this response can consist of either the provision or elicitation of linguistic information (Ellis et al., 2001a, 2001b; Lyster, 1998a, 1998b; Lyster & Ranta, 1997). Within these broad response categories are specific types of response moves. For example, the provision of linguistic information can occur through explicit correction or recasts (i.e., reformulation of the learner's erroneous utterance). The second response type is elicitation (which Lyster, 1998a, terms "negotiation of form"), which can occur through clarification requests, repetition, and direct requests. Both Lyster's and Ellis et al.'s studies found that the most frequent response type was the provision of information, particularly through recasts, although Lyster (1998a) noted that the type of response move varied according to the linguistic focus of the learners' errors. Grammatical and phonological errors received more recasts, whereas lexical errors received more negotiation of form (i.e., elicitation). Furthermore, a recent review of recasts (Nicholas, Lightbown, & Spada, 2001) suggested that the efficacy of recasts might be influenced by the targeted linguistic feature, the amount of attention brought to bear through the task and context, and the developmental readiness of the learner, among other factors.

After the response move, it is possible for learners to produce uptake, which Lyster and Ranta (1997) defined as "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the participant's initial utterance" (p. 49). The production of uptake is optional, as it does not necessarily occur after the provision of feedback; furthermore, if the teacher continues his or her turn, learners might not always have the opportunity to produce uptake (Lyster, 1998a; Oliver, 2000). Additionally, it has been noted that not all uptake is of the same quality. Lyster and Ranta categorized uptake as *repair* or *needs-repair*, according to whether or not learners successfully corrected their erroneous utterances. Ellis et al. (2001a), building on Lyster and Ranta's definition, distinguished between successful and unsuccessful uptake, a distinction that depended on whether or not participants incorporated the linguistic information targeted in the teacher's response into their own production. Studies have found varying levels of successful uptake, including 27% (Lyster & Ranta), approximately 35% (Oliver, 1995), 74% (Ellis et al.), and 47–75% (Loewen, 2004). These studies acknowledged that uptake is not necessarily indicative of L2 acquisition; nevertheless, uptake might provide some indication regarding the internal processes of L2 learners. As Lightbown (1998) suggested, "a reformulated utterance from the learner gives some

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

reason to believe that the mismatch between learner utterance and target utterance has been noticed, a step at least toward acquisition" (p. 193).

In addition to identifying discourse moves such as response and uptake, researchers have analyzed incidental focus on form for a variety of psycholinguistically motivated characteristics (Ellis, Loewen, & Basturkmen, 1999). To date, such studies have been primarily descriptive, but these analyses can serve as a basis for investigating which characteristics, if any, might contribute to the effectiveness of incidental focus on form.

The cause or source of incidental focus on form—whether it occurs as a result of miscommunication or simply from a desire to focus on the accurate use of a linguistic structure—is one such characteristic that has been investigated (Ellis et al., 2001a, 2001b). The interaction hypothesis (Long, 1996) suggests that negotiation of meaning following communication breakdown will benefit learners. However, researchers also acknowledge that negotiation of form can occur in meaning-focused classes, even if there is no breakdown in communication (Pica, 2002). For example, Ellis et al. (2001a) found that negotiation resulting from a purely linguistic problem was twice as frequent as negotiation resulting from a communication problem in the meaning-focused lessons they observed. Nevertheless, uptake was significantly more likely to occur in FFEs that involved negotiation of meaning.

Other characteristics of incidental focus on form may relate to its potential to induce noticing of targeted structures. In particular, Schmidt (1990, 1995, 2001) argued that factors such as salience of the linguistic items can affect whether learners will notice those items. Thus, the directness of the response move might influence learners' tendency to notice linguistic items, with more indirect responses—such as recasts—being less noticeable (Lyster, 1998a). Similarly, the complexity or length of negotiation surrounding a linguistic item might influence the saliency of that item, with longer negotiation sequences being potentially more salient. Ellis et al. (2001a) found that complex FFEs were significantly more likely to contain both uptake and successful uptake.

Effectiveness of Incidental Focus on Form

Even though previous studies have described the occurrence of incidental focus on form and uptake, they have not been able to address the relationship between incidental focus on form and L2 learning. This raises the question of how to measure the effectiveness of incidental focus on form. With planned focus on form, it is possible to use a pretest and posttest to measure gains in learners' ability to use the targeted structure(s); however, with incidental focus on form, such pretesting is not feasible because of the unpredictability of the linguistic items that arise during a meaning-focused activity (Swain, 2001). Perhaps the best measure of the effectiveness of incidental focus on form would be learners' subsequent spontaneous production of the targeted linguistic forms; however, such data are often difficult to obtain (Wil-

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

liams, 2001). Additionally, although subsequent correct production could be considered an indication of the effectiveness of a FFE, a lack of production is not necessarily evidence of focus on form's ineffectiveness. Rather, learners might simply not have had occasion to use the linguistic form in question. One method for circumventing these difficulties is to elicit production of the targeted linguistic items through the use of individualized, tailor-made tests based on the specific linguistic items that arose in the classroom discourse (Swain, 1995, 2001).

Individualized testing allows researchers to assess the specific linguistic items targeted in spontaneous FFEs in classroom discourse. However, because the testing can be done only after the incidental focus on form has occurred, it is not possible to investigate the learner's prior knowledge of the linguistic item. Thus, it is possible that the error is a performance error. Nevertheless, the fact that an error in production has occurred or that a question about a linguistic item has been raised indicates learner difficulty with that item (Ellis et al., 2001b). The fact that learners cannot use a form correctly in meaning-focused interaction—as evidenced by the FFE—indicates that further consolidation of learning might need to occur (Swain, 2001). Thus, learning can be operationalized as an increase in the accurate use of the targeted forms in subsequently elicited situations (Williams, 2001).

Several recent studies have begun to explore the relationship between incidental focus on form and L2 learning by individually testing the linguistic items that have arisen incidentally. For example, Swain (1995) discussed a study by La Pierre (1994) that investigated learner negotiation about language form in a text reconstruction activity. La Pierre found that learners were 80% accurate in a 1-week-delayed posttest on tests of the individual items. Williams (2001), in a study of eight learners' spontaneous attention to form in an intensive English program, found that learners' accurate performance on individualized tests ranged from 45% to 94%, with higher-proficiency learners scoring higher. Finally, Nabei and Swain (2001) used grammaticality judgment tests to investigate the ability of one Japanese-speaking learner of English as a foreign language to benefit from recasts provided during 7 hours of an English discussion course. They found an overall accuracy rate of 56% on the 27 items tested.[1] Although these studies have started to examine the effectiveness of incidental focus on form through individualized testing of the linguistic items, they have been fairly limited in scope. The present study attempts to extend this research by asking the following research questions:

1. What effect, if any, does incidental focus on form have on learners' L2 learning?
2. What characteristics of incidental focus on form, if any, predict L2 learning?

METHOD

This research is part of a larger study investigating the occurrence and effectiveness of incidental focus on form (Loewen, 2002, 2003, 2004). The current

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

paper examines the relationship between incidental focus on form and L2 learning, as operationalized by the individualized testing.

Participants

The research was conducted at a private language school in Auckland, New Zealand. A total of 12 intact classes with 12 different teachers and 118 learners participated in the study. All 12 teachers were native speakers of English. They had between 1 and 15 years of teaching experience and held a variety of English-as-a-second-language qualifications, such as the Certificate in English Language Teaching for Adults and Master of Arts in Applied Linguistics. Eight of the teachers were male and four were female. The only demographic information gathered directly from the learners was their nationality, with the majority (over 75%) coming from Korea, China, Japan, and Taiwan. Females accounted for 56% of the participants and males accounted for 44%. The classes ranged in attendance from 3 to 14 learners. The English language proficiency of the learners, as measured by an in-house placement test, ranged from low to upper intermediate.

Procedure

A total of four lessons were observed over a 1-week period for each of the 12 classes. The researcher was present during all observations as a nonparticipant observer, and the observations were audio-recorded using a wireless cassette recorder with a clip-on microphone attached to the teacher. This arrangement recorded all teacher-learner interaction, whether as a whole class, in small groups, or one-on-one; however, it did not capture learner-learner interaction when the teacher was not present. Because it was important to identify individual learners' utterances on the recordings, detailed notes consisting of learners' names and speech excerpts were made during the observations and matched to the recordings. A total of 32 hours (1,917 minutes) of meaning-focused classroom interaction were observed and recorded. However, in an effort to minimize the disruption of the testing procedures, only the data gathered from the second and third observations were used for the individualized testing, although the first and fourth observations were used to supplement the description of the occurrence of incidental focus on form in these classes. Thus, a subset of 17 hours (1,022 minutes) of data was used for testing. (The testing procedures will be described in detail in the following section.)

Although instruction at this language school consisted of both meaning-focused and form-focused lessons, the present study focused exclusively on meaning-focused activities, which were defined as those activities having a primary goal of exchanging information, rather than learning about or practicing specific linguistic forms (Pica, Kanagy, & Falodun, 1993). This definition, however, was not provided to the teachers and no effort was made to guide

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

them in their choice of lesson plans beyond asking them to observe meaning-focused lessons. Activities observed included information- and opinion-gap tasks, role plays, and class discussions on a variety of topics. Because the study focused only on the four observed lessons, it is not possible to comment on the nature or content of lessons that were not observed, and it is acknowledged that the observed lessons comprised only a small part of the learners' instruction.

Identification and Testing of Focus on Form Episodes

Following the observations, the FFEs were identified. A FFE was defined as consisting of "the discourse from the point where the attention to linguistic form starts to the point where it ends, due to a change in topic back to message or sometimes another focus on form" (Ellis et al., 2001a, p. 294). Thus, in practice, a FFE started either when a learner produced a linguistic error that was addressed by the teacher or when a learner queried a linguistic item.

(2) L: so *he is in prison for six years*
T: *he has=*
L: *=has been in prison for six years*
T: yes
L: I don't know how long he should be

An example of a FFE is shown in (2). The learner and teacher are discussing which of four prisoners should be paroled early (an information- and opinion-gap task). The FFE (which is italicized; see the Appendix for a full transcription key) begins when the learner makes an error in the verb form, which is immediately corrected by the teacher. The learner repeats the correct form, and they continue discussing the prisoner's case. Thus, the participants' attention has switched briefly from meaning-exchange to a linguistic item and then returned to the main message. To determine reliability of FFE identification, a second rater coded a sample of 10.3% of the data (three lessons totaling 199 minutes), with a resulting agreement rate of 89.9%.

Because it was not possible to predict which linguistic items would be targeted in a particular lesson, the FFEs had to serve as a type of pretest, demonstrating—either through an error or a query—that a learner had difficulty with a particular linguistic structure. Once the FFEs were identified, individualized test items relating to the linguistic items targeted in the FFEs were constructed, and the learners who were responsible for triggering the specific FFEs were tested on those items. Tests were given on two different occasions: one immediate (1 to 3 days after the FFE) and one delayed (13 to 15 days after the FFE). However, because a pilot study suggested that the tests served to focus attention on the linguistic forms, it was not possible to conduct immediate and delayed tests on the same linguistic items. Therefore, immediate testing was done on linguistic items from the second observation and delayed

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

testing on items from the third observation. Thus, for each class, only linguistic items from two of the four observations were used for testing. Additionally, the immediate testing (based on the second observation) was conducted after the third observation (which provided the data for the delayed test) so that the testing would not heighten learners' awareness of subsequent FFEs and thus affect their responses to the delayed testing. Finally, note that the linguistic content of instruction between the observations and the delayed testing was not documented; although the incidental and extensive nature of the linguistic structures targeted in the FFEs suggests that the majority of structures would not have been attended to, it is possible that some structures might have been focused on in subsequent lessons.

For the administration of the tests, learners were withdrawn individually from class into a quiet room where the researcher presented the test items to them. The test items were based as closely as possible on the FFEs. However, to facilitate testing, three templates with slightly differing formats—suppliance, correction, and pronunciation—were developed. For suppliance and correction tests, the researcher read aloud the test items to the learners. To replicate as closely as possible the oral nature of the FFEs, learners were not given written copies of the test items. Each test item was read to the learner no more than twice; however, learners could respond after the initial presentation if they wished. For pronunciation tests, students were asked to read aloud the test items, because of the difficulty of eliciting the targeted items without providing written prompts. The learners responded orally to all test items, and each testing session was audio-recorded.

Suppliance. Suppliance tests were used primarily for FFEs targeting vocabulary and spelling and required learners to provide linguistic information. For example, the test rubric in (3), which the researcher read aloud to the learner, asks the learners to supply the meaning of *increasing*. The test item was based on episode 456, presented in (4), in which the learner asked about the meaning of this word during a discussion activity.

(3) Please tell me the meaning of the following word: Increasing

(4) L1: *excuse me what does this word mean?* (pointing to *increasing* in the text)
T: *increasing* (.) *growing*
L1: *growing*
T: *growing yeah* (.) *so today there are more and more smoke free* (.) *no smoking cafes* (.) *more cafes* (.) *cinemas say no smoking*
L2: *yeah*
T: *everyday more and more and more* (.) do you like this?

Suppliance tests could also ask learners to provide a word or phrase that corresponded to a particular meaning or definition, such as the test rubric in (5), in which the learner needed to supply the response *addicted*. To reduce the chance of learners providing an acceptable answer other than the target

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

of the FFE, learners were told the first letter of the target word (Laufer & Nation, 1999). The FFE (episode 462) on which this test item was based is shown in (6).

(5) When someone has to have drugs, they can't stop, we say they are _____. It begins with the letter A.

(6) L1: *how can I say uh maybe if I had have drug*
T: *mhm*
L1: *have a drug* (.) *and I can't stop*
T: *addicted*
L1: *addicted*
T: *addicted* (.) *I am addicted*
L1: *ah*
T: *yeah* (.) *I am addicted* (.) *you can be addicted to exercise* (.) *addicted to chocolate* (.) *you can't stop* (.) *okay*
L2: addicted to drugs
T: addicted to drugs (.) yeah you can't stop (.) can't stop (.) okay

Correction. In correction tests, learners were asked to improve sentences that they had produced incorrectly during FFEs. For example, in the test rubric in (7), a learner was given the opportunity to correct his preposition error, which was the focus of the FFE, episode 648, shown in (8).

(7) The following sentences are incorrect or inappropriate. Please listen and tell me how you could make the sentences better: I used to wear the balaclava for protection to wind and cold.

(8) L: when I was soldier I used to wear the balaclava
T: and why did you wear it (learner name) for protection from the cold or for another reason
L: just wind uh *protection to wind and cold*
T: *protection from*
L: *uh from wind and cold*
T: right, okay not for a disguise

Pronunciation. Pronunciation tests were the only tests that involved written prompts. Learners were first asked to read aloud a sentence containing the targeted word or phrase and then to read aloud the word or phrase in isolation. Although the written form of the word provided learners with clues about pronunciation, it was decided that this was an unavoidable limitation because of the difficulty of eliciting the targeted items without some type of prompt. Thus, the learner's mispronunciation of *scalp* in episode 400, shown in (10), is tested by the test rubric in (9) that asks him to read aloud first his original utterance and then the word in isolation.

(9) Please read aloud the following sentence and word:
The title of the story is girl has bullet in her scalp.
Scalp

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

(10) L1: *the title of the story is girl has bullet in her scamp*
T: *scalp*
L1: *scalp*
L2: scalp
T: *scalp* (.) *this bit*
L2: *scalp yep*

Because learners were tested only on the FFEs in which they participated directly (either by making an error or raising a query), the number of test items for each learner varied. Some learners had no test items on a particular day, whereas the highest number of test items was 14. (See Loewen, 2003, for a further description of the variation in learner participation in incidental focus on form.)

Scoring of Test Items

Learners' responses to the test items were scored as correct, incorrect, partially correct, other correct, assisted correct, or inconclusive.

Correct. The learner produced a response that correctly matched the targeted linguistic item in the FFE. For example, in response to the test item in (11), the learner provides the correct form, *depressed*, which was based on episode 612, shown in (12).

(11) R: okay so the sentences are incorrect and I'd like you to make them better.
L: okay.
R: The first one (.) he was depression in prison.
R: he was depression in prison.
L: um (.) depress (.) depressed.
R: alright (.) good.

(12) L: *he was depression in prison because*
T: *oo this is what part of language is this*
L: *depressed*
T: *depressed great excellent well done*
L: he was depressed in prison because he wife divorced with him and take the two sons away from he

Incorrect. The learner does not correctly produce the linguistic item targeted in the FFE. The example in (13) illustrates an incorrect response, *alcoholer*, as well as the response that would be expected, *alcoholic*. This example is based on episode 618, which is presented in (14).

(13) R: what do we call people who like alcohol?
L: alco- alcoholer
Expected response: alcoholic

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

(14) L: the people who like alcohol (.) *alcoholist*
T: *alcoholic*
L: *alcoholic*
T: good okay

Partially Correct. The learner produces a response that improves on the targeted linguistic error in some way but is still not entirely accurate. This rating was also given to pronunciation test items in which one context (e.g., the sentence) was incorrect, but the other context (e.g., isolated) was correct. For example, in (15) [based on episode 585, which is presented in (16)], the learner mispronounces *nephews* in the sentence but pronounces it correctly in isolation. Thus, the test item is scored as partially correct.

(15) R: Please read aloud the following sentence
L: She has many niece and nephees.
R: Please read aloud the following word
L: nephews

(16) L: *she has many niece and nev*
T: *nephews*
L: *nephews*
T: *so that's* (.) *wo wo can we just go back* (.) *what's this*
L: *niece*
T: *nieces*
L: *nieces*
L2: *nieces*
T: *and*
L3: *nephews*
L: *and nephews*
T: *nephews* (.) yeah

Other Correct. The learner corrects the targeted error, but in a way that differs from the linguistic information provided in the FFE. Thus, in the example in (17), the learner has provided the arguably acceptable word *attached* instead of *assigned*, which was provided in episode 138 [shown in (18)].

(17) R: The government A_____ a policeman to the school
L: the government attached a policeman to the school

(18) L: *the government hatched policeman to this*
T: *has what? the government*
L: *prohatch-ed*
T: *patched*? *um uh assigned*
L: *assign*
T: *assigned* [*would be a good word*
L: [*assigned yeah*
T: mm

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Assisted Correct. Interviewer accommodation can be an issue in face-to-face, oral testing (Ross, 1992) and, in spite of the best efforts of the researcher, the transcripts of the testing revealed several occasions when participants received additional help in producing a response (e.g., the researcher repeated the sentence more than twice or pinpointed the error). These test items were excluded from the analysis.

Inconclusive. Occasionally, it was not possible to make a judgment about the correctness of a response, because of factors such as indecipherable learner utterances and the nonsalience of word final morphemes. These responses were excluded from the analysis.

Test Reliability and Validity

With any type of language testing, it is important to establish the reliability and validity of the tests (Douglas, 2001). However, with this type of individualized testing, it was not possible to use traditional means of establishing reliability, such as test-retest or internal consistency. Consequently, an alternative approach to establishing reliability was taken: Potential sources of error variance or measurement error were identified (Brown, 1996) and an attempt was made to neutralize them. (See Loewen, 2002, for a complete description of the efforts to ensure reliability.) However, the tests still comprised different linguistic information and different numbers of items. Thus, it was decided that the best way to determine the suitability of the test items was to judge their construct validity, particularly given that reliability is often seen as a component of validity (Douglas).

Construct validity refers to "the extent to which we can interpret a given test score as an indicator of the ability(ies) or construct(s) we want to measure" (Bachman & Palmer, 1996). The aim of the present study was to investigate to what extent the test items measured learners' ability to produce the linguistic information they were provided in FFEs. Thus, one way to establish construct validity was to determine how well the test items related to what occurred in the FFE. Although the specific linguistic items targeted in the FFEs were varied and unpredictable, the FFEs, nevertheless, had certain characteristics in common. Thus, the validity of the test templates—suppliance, correction, pronunciation—was investigated in order to determine if they could be used satisfactorily on a variety of linguistic items. With that purpose in mind, a panel of three experts in the field of SLA and L2 testing was convened. The panel members were provided with transcripts of a random sample of approximately 10% of the tested FFEs. The panel members then listened to the audio-recorded test items and were asked to rate on a 4-point scale (1 corresponded to highly appropriate, whereas 4 corresponded to inappropriate) how well the items tested the linguistic information targeted in the FFEs. The results from this rating session showed that the average rating for the suppliance and pronunciation test items was highly appropriate (mean = 1.17) and that there was

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

considerable agreement (94.7%) among the three raters. However, the correction test items did not have similarly high ratings or agreement. It was therefore decided to conduct a second rating session in which all of the correction test items ($n = 149$), plus several other test items that concerned the researcher ($n = 19$), were rated by the same three raters. A total of 21 test items had means over 2.5, which was considered to indicate low construct validity; these test items were withdrawn from the analysis. Reasons given by the panel for low validity scores included the decontextualized nature of the test items in relation to the highly contextualized nature of the FFE interaction as well as the inconclusiveness of some FFEs in which negotiation of linguistic forms was never fully resolved. Finally, the reliability of the coding of the test responses as correct, incorrect, and so forth was confirmed by a second rater, who scored a random sample of 21% of the test items. This resulted in a reliability level of $\kappa = .850$, which is above the .80 level suggested by Davies et al. (1999).

Coding of FFEs

After the FFEs were identified and tested, they were transcribed in detail and coded for a variety of characteristics, based primarily on the analysis in Ellis et al. (1999). Table 1 identifies and defines these characteristics.

An example shown in Table 2 will help illustrate the coding system. In episode 560, presented in (19), the participants are discussing which of four candidates should receive a heart transplant (an information- and opinion-gap activity). The learner makes a grammatical error in the formation of the plural *childrens*, which the teacher recasts. The learner incorporates this feedback successfully into his production; however, the teacher provides another piece of linguistic information by informing the learner of what is wrong with his utterance (*no S*). Thus, the episode is complex and heavy because the teacher provides two different response moves. The FFE is code-related because there is no apparent breakdown in communication.

(19) L: she works full time and uh *pay take care of her childrens until she* ⟨*the time she come back*⟩
T: *children*
L: *children*
T: *no S*
L: *yeah children* (.) and she get no finance finance <finance-al> help . . .

To investigate the reliability of the coding, a second rater coded a sample of 21% of the FFEs ($n = 277$). Table 3 lists the kappa reliability and percentage agreement scores for each of the characteristics. The kappa coefficients ranged from $\kappa = .672$ to .915, and all percentage agreement scores were over 85%.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Table 1. Characteristics of FFEs

Characteristic	Definition	Categories
Type	Instigation	*Reactive*: Error correction *Student-initiated*: Query raised by L
Linguistic focus	Linguistic target	*Grammar* *Vocabulary* *Pronunciation*
Source	Apparent reason for instigation	*Code*: Inaccurate use of linguistic item with no apparent miscommunication *Message*: Problem understanding meaning
Complexity	Length	*Simple*: Only one response move *Complex*: More than one response move
Directness	Explicitness of the feedback	*Indirect*: Implicit (e.g., recast) *Direct*: Explicit (e.g., metalingual explanation)
Emphasis	Complexity + Directness	*Light*: Indirect and Simple *Heavy*: Direct, Complex, or both
Timing	Response timing	*Immediate* *Deferred*
Response	Type of feedback provided by teacher	*Provide*: T gives information about a language form either by use of a recast or an inform *Elicit*: T attempts to draw out from L(s) a language form or information about a language form
Uptake	Student response to feedback	*Uptake*: L produces response *No uptake*: L does not respond *No opportunity*: L does not have a chance to respond
Successful uptake	Quality of student response	*Successful uptake*: L incorporates linguistic information into production *Unsuccessful uptake*: L does not incorporate linguistic information into production

Statistics

Raw frequencies and percentages were calculated for the test response categories. Because the data consisted of frequency counts of categorical data, Pearson's chi-square analysis was performed on the raw frequencies. An alpha level of $p < .05$ was set for all chi-squares. Because analyses involved variables with more than two coding categories (i.e., the chi-square was larger than 2×2), adjusted standardized residuals of greater than plus or minus 2.0 were used to identify significant differences. All inferential statistics were performed using SPSS 10.0.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Table 2. Example of coding scheme

	Episode 560
Characteristics	Category
Type	Reactive
Linguistic focus	Grammar
Source	Code
Complexity	Complex
Directness	Direct
Emphasis	Heavy
Response	Provide (inform)
Timing	Immediate
Uptake	Uptake, Successful

To investigate causality in the data, a binary logistic regression analysis was performed (Hatch & Lazaraton, 1991; Saito, 1999). Logistic regression is similar to linear regression; however, it is distinguished by the fact that the dependent variable is binary or dichotomous rather than continuous (Greenhouse, Bromberg, & Fromm, 1995). The purpose of logistic regression is to find the "best fitting and most parsimonious" model to describe the relationship between a dependent variable and independent (or predictor) variables (Hosmer & Lemeshow, 2000). In the case of the present data, it was desirable to use regression modeling to determine which characteristics of incidental focus on form would best predict the test scores. Because the dependent variables were categorical rather than continuous, logistic regression was appropriate. Finally, because of the slightly different structures of the suppliance, correction, and pronunciation tests, it was decided to conduct separate regressions for each template.

Table 3. Reliability scores

Characteristic	κ	Percentage agreement
Type	.915	96.4
Linguistic focus	.835	88.8
Source	.684	85.2
Complexity	.747	87.4
Directness	.753	87.7
Emphasis	.830	92.4
Response	.672	94.2
Timing	.745	97.4
Uptake	.859	95.3
Successful uptake	.899	95.9

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Given that logistic regression allows only binary dependent variables, it was necessary to dichotomize the dependent variable (in this case, the test scores). Because both partially correct and other correct responses contained an element of correctness, it was decided to recode them as correct. This recoding increased the sample size and an exploratory analysis with and without such recoding produced nearly identical results. Furthermore, the assigning of binary values to the independent variables makes the interpretation of the resulting odds ratios (which will be discussed later in this section) easier. Table 4 details the coding scheme for the dependent and independent variables and indicates the value assigned to each coding category.

The method of entry for each variable in the regression equation was forward stepwise, which involves adding the independent variables to the model one at a time in an attempt to produce the best and most succinct model. Because each step adds the variable that makes the greatest change to the model, stepwise regression is particularly appropriate when the relationships between the predictor variables and the dependent variable have not been previously investigated or are relatively unknown. If a variable does not make a significant contribution to the model, it is not included (Hosmer & Lemeshow, 2000). Thus, a forward stepwise regression produces a model that accounts for the largest proportion of the dependent variable using the fewest—but most important—independent variables. Hosmer and Lemeshow argued that setting an alpha level of .05 for a forward stepwise regression is too stringent and might exclude important variables; instead, they "highly recommend" choosing an alpha level in the range of .15 to .20 (p. 118). For the purposes of the current study, an alpha level of .15 was chosen.

The output of a logistic regression analysis includes odds ratios (ORs) and 95% confidence intervals for each independent variable. ORs are important

Table 4. Binary coding scheme

	Value	
Variable	0	1
Test score	Incorrect	Correct
Type	Reactive	Student-Initiated
Linguistic focus[a]	—	—
Source	Code	Message
Complexity	Simple	Complex
Directness	Direct	Indirect
Emphasis	Light	Heavy
Response	Provide	Elicit
Timing	Immediate	Deferred
Uptake	No uptake	Uptake
Successful uptake	Unsuccessful uptake	Successful uptake

[a]Not reducible to a binary distinction.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

because of their relationship to the regression coefficient (OR = $e^{\beta 1}$) and their ease of interpretation. "The odds ratio then approximates how much more likely (or unlikely) it is for the outcome to be present among those with x = 1 than among those with x = 0" (Hosmer & Lemeshow, 2000, p. 50). For example, if the dependent variable of test score (coded as 0 = incorrect and 1 = correct) and the independent variable of complexity (coded as 0 = simple and 1 = complex) have an OR of 2, then the OR estimates that correct test scores are twice as likely to occur when based on complex FFEs than on simple ones in the dataset. Conversely, an OR of 0.5 estimates that correct test scores are one-half as likely to occur when based on complex FFEs than on simple FFEs in the dataset.

Finally, it should be mentioned that one assumption of chi-square and logistic regression analyses is that the data are independent, meaning that "each observed frequency is generated by a different subject" (Gravetter & Wallnau, 1991). This assumption, often violated in SLA research, can lead to Type I errors (i.e., the overestimation of significant differences in the data; Hatch & Lazaraton, 1991; Saito, 1999). In the present study, the test scores are not entirely independent because multiple scores are produced by individual learners (i.e., individual learners often had more than one test item). To test for the effect of this arguably limited dependence of data, a participant variable for learners was created and included in the logistic regression models, as suggested by Saito. The results indicated that although the learner variable was a significant predictor of test scores in the correction and pronunciation tests, the ORs and significance levels for the other independent variables entered into the model did not change substantially when the participant variable was excluded. These results suggest that although the data are not completely independent, any dependence in the data does not substantially affect the regression models. (See Loewen, 2002, for further discussion of this issue.)

RESULTS

A total of 473 FFEs were tested; approximately one-half were tested 1 day after the FFE (immediate) and the other half were tested 2 weeks after the FFE (delayed). Table 5 shows the response scores on the immediate and delayed tests. On the immediate test, 47.6% of the responses to test items were correct, whereas in the delayed test, 39.3% were correct. There were more partially correct answers in the immediate testing (14.8% vs. 9.8%), whereas other correct answers were more frequent in the delayed testing (2.6% vs. 8.3%). Assisted correct and inconclusive responses accounted for approximately 4% of the immediate and delayed testing, and these items have been excluded from the rest of the analyses in this section.

A chi-square analysis was performed to test for statistically significant differences in the frequency of correct responses in the immediate and delayed testing. The results were significant, χ^2 (3, 452) = 12.48, $p < .05$. The residuals

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Table 5. Test scores

	Immediate		Delayed		Total	
Test score	*n*	%	*n*	%	*n*	%
Correct	109	47.6	96	39.3	205	43.3
Incorrect	71	31.0	92	37.7	163	34.5
Partially	34	14.8	24	9.8	58	12.3
Other correct	6	2.6	20[a]	8.2	26	5.5
Assisted correct	6	2.6	9	3.7	15	3.2
Inconclusive	3	1.3	3	1.2	6	1.3
Total	229		244		473	

[a]Adjusted standardized residual > 2.0.

indicate that other correct responses occurred more frequently in the delayed testing.[2]

The three regression analyses—with the dependent variable of test scores and independent variables of type, linguistic focus, source, complexity, directness, emphasis, response, timing, uptake, and successful uptake—are presented in Table 6. For the correction test, the overall percentage accuracy was 64.2%. Table 6 details the two predictor variables entered into the model. The most significant variable was successful uptake, which had an OR of 2.1, meaning that correct responses were twice as likely when the FFE on which the test item was based contained successful uptake rather than unsuccessful uptake. The second and only other variable included in the model had a significance level of .146, meaning that it barely passed the exploratory alpha level of .15 for entry into the model. The OR indicates that FFEs with uptake were half as likely to result in correct scores.

Table 6. Logistic regression

		Confidence intervals[a]			
Task	Predictor variables	Lower	Upper	Odds ratio	α
Correction	Uptake	0.217	1.255	0.522	.146
	Successful uptake	1.063	4.344	2.149	.033
Suppliance	Uptake	3.263	23.716	8.797	.000
	Successful uptake	0.063	0.692	0.208	.010
Pronunciation	Complexity	0.133	0.722	0.309	.007
	Source	1.184	7.582	2.997	.020
	Successful uptake	0.832	4.369	1.906	.127

[a]95%.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

The regression analysis for the suppliance test is similar to the correction analysis. The overall percentage correct was 74.8%, and Table 6 details the two predictor variables entered in the model. Again, successful uptake was the strongest predictor variable, with an OR of 8.7, meaning that FFEs with successful uptake were almost nine times more likely to result in correct scores on the test items than FFEs without successful uptake. Furthermore, FFEs with uptake were one-fifth as likely to result in correct test scores.

The pronunciation regression produced a somewhat different model from the previous two. The model had an overall percentage correct of 66.9%. Table 6 presents the details of the three predictor variables entered in the model. Complexity had an OR of 0.309, meaning that complex FFEs were one-third less likely to result in correct scores than were simple FFEs. Source had an OR of 2.9, meaning that message-related FFEs were almost three times more likely to result in correct test scores than were code-related FFEs. Finally, successful uptake had an OR of 1.9, meaning that FFEs with successful uptake were almost twice as likely to result in correct scores than were FFEs with unsuccessful uptake.

DISCUSSION

These results suggest that the answer to the first research question is that incidental focus on form does have some effect on L2 learning, as measured by the individualized tests. Participants were able to remember and produce the targeted linguistic items almost 50% of the time in the immediate testing context and almost 40% in the delayed context. If partially correct answers are considered as well, the correct response rate rises to 62% for the immediate testing and 50% for the delayed. These results are roughly comparable to other studies (La Pierre, 1994; Nabei & Swain, 2001; Williams, 2001), which have found participants performing accurately on similar individualized tests. Although the scores might not seem remarkably high, it should be remembered that the targeted linguistic items were addressed incidentally in lessons that were not specifically designed to address these items. Thus, the fact that participants were able to supply the correct linguistic information in at least half of the FFEs in which they participated is encouraging, given the incidental and generally brief nature of the FFEs.

Of course, without the possibility of a pretest because of the unpredictability of the targeted linguistic items (Swain, 2001), the interpretation of the tests is somewhat difficult. In this study, the occurrence of a FFE was taken as an indication that the learner had incomplete knowledge of a linguistic item, because the learner either made an error in the use of that item or asked a question about it. However, it remains unclear whether the error or query represented a gap in the learner's knowledge or an inability to produce a known structure in oral production. Therefore, it is not possible to determine if correct test scores represent the incorporation of new linguistic knowledge into

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

the learner's interlanguage system or if they represent a consolidation of previously existing knowledge. If this issue is phrased in terms of learner's implicit and explicit L2 knowledge (Bialystok, 1994; Ellis, 2004), it is possible that the FFE represents a lack of implicit, proceduralized knowledge on the part of the learner, which prevents accurate production during relatively spontaneous oral production. However, the FFE might not provide information about the learner's explicit, declarative knowledge of the structure, because the learner might not have had time to access that knowledge. However, because of the unlimited time allowed the learners, the decontextualized nature of the test items, and the full attention on form rather than on both meaning and form, the testing might better reflect learners' explicit knowledge. Further investigation that measures learners' ability to use linguistic information from FFEs in more spontaneous production would be useful, although this method is not without its difficulties (Williams, 2001).

With regard to the second research question, the results of the logistic regressions show that among the various characteristics of incidental focus on form, successful uptake was the characteristic primarily associated with correct test scores. The regression models for the correction and suppliance tests were very similar and found that successful uptake was the best predictor of correct scores on the test, with the effect being stronger for the suppliance tests. However, when the effects of successful uptake were taken into consideration, there was still a slight effect for uptake: The analysis showed that an absence of uptake was slightly related to correct test scores. This result might mean that simple acknowledgment of the teacher's feedback or repetition of an error was worse than not responding at all. These results would suggest that it is not simply the presence of uptake but rather the quality of uptake that is important.

Although successful uptake appeared in the regression analysis for pronunciation test items, the results showed that simple FFEs and message-related FFEs also predicted correct test scores. The results suggest that different kinds of focus on form might be needed for different aspects of language (Nicholas et al., 2001); indeed, other research has found that patterns of uptake can differ according to the linguistic focus of the error and the type of feedback provided (Lyster, 1998a; Oliver, 1995). A possible interpretation for the current results might be that pronunciation items are less amenable to incidental intervention. If learners make pronunciation errors and they produce the correct form in one turn, then perhaps they have some control over that phonological feature. However, if they are not able to correct the item on the first turn, then the extra effort of complex feedback might prove ineffective. Similarly, with message-related FFEs, it might be that phonological problems that interfered with communication were more within learners' ability to improve than were the pronunciation items that were focused on solely for accuracy's sake. However, these explanations are speculative and warrant further investigation.

The inclusion of source in the pronunciation regression model—but not in the other two—raises questions about the effectiveness of meaning negotia-

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

tion. Although message-related FFEs successfully predicted learners' correct scores on the pronunciation tests, there was no difference between code- and message-related FFEs on the suppliance and correction tests. This result suggests that although negotiation of meaning positively affected learners' pronunciation, it was no more effective than negotiation of code in helping learners retain the targeted vocabulary and grammar items. It might be that negotiation is helpful for learners, regardless of whether it results from miscommunication or from attention to linguistic accuracy for its own sake. It could also be that the benefit of negotiation of code occurs because the targeted linguistic items are more readily apparent than they are in negotiation of meaning. Indeed, Oliver (1995) found differing patterns of feedback depending on the clarity of the original utterance. If a learner's intended meaning is readily apparent—as in code-related FFEs—the teacher can supply or elicit precise linguistic information appropriate to the FFE, whereas in message-related FFEs, the intended linguistic forms can often be unclear. In a number of FFEs in the present study, the learner's intended meaning was never made clear; consequently, there was no readily identifiable linguistic information for the learner to attend to.

One theoretical implication of this study concerns the role of output. The results of this study tend to support Swain's (1985, 1995, 2000) output hypothesis, which argues that pushed output forces learners to process syntactically rather than semantically and that this processing encourages greater linguistic accuracy. In this study, successful uptake was found to be a predictor variable in each of the three regression models examining the test scores. The results indicate that it was not merely the presence of uptake that was important; in fact, uptake that was incorrect or still in need of repair was apparently not helpful. Learners needed to actually produce the correct linguistic items for there to be an improvement in test scores.

It should be noted, however, that these results do not clearly specify the role of successful uptake in L2 learning. It might be that learners were more able to produce successful uptake and, consequently, accurate test scores, because the targeted linguistic items were already partially known to them. Therefore, the attention to form allowed them to draw on this knowledge; thus, the production of successful uptake resulted from, rather than promoted, the consolidation of L2 knowledge. Similarly, it could be that noticing—rather than producing—linguistic items is impacting the interlanguage system. From the current data, it is difficult to determine the precise role of successful uptake. To determine if output was indeed the important feature for these learners, it would be necessary to compare linguistic items that they noticed but did not produce with items that they noticed and for which they produced successful uptake. Such an investigation might be conducted in a more controlled, experimental-type study or perhaps through stimulated recall (Gass & Mackey, 2000).

In conclusion, the results of this study suggest that incidental focus on form helps learners improve their linguistic accuracy while they are engaged in

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

meaning-focused L2 lessons. Additionally, the regression analyses found that successful uptake was the best overall predictor of test performance, suggesting that it might be important for learners to actually produce the targeted linguistic items. However, the limitations regarding the interpretation of the results suggest that further research is needed, particularly into (a) the nature of the linguistic knowledge represented in FFEs, (b) the most effective ways of measuring any learning resulting from FFEs, and (c) the precise role of successful uptake in the learning process.

(*Received 12 September 2004*)

NOTES

1. Nabei and Swain (2001) used a delayed test in which all of the previous items were repeated; however, in the interval between the first test and the second test, the learner was asked to comment on the recast episodes through the use of stimulated recall. The authors admit that, as a result, the second test does not provide a measure of the effect of the recast by itself, but also of the opportunity to reexperience it and talk about it. As such, the results from the second test are not discussed in the present paper.

2. The residual for correct responses was 1.7, which is close to the 2.0 level, indicating a trend for more correct responses in the immediate testing.

REFERENCES

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*, 157–168.

Brown, J. D. (1996). *Testing in language programs*. Englewood Cliffs, NJ: Prentice Hall.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: University of Cambridge Local Examinations Syndicate.

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

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

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.

Douglas, D. (2001). Performance consistency in second language acquisition and language testing research: A conceptual gap. *Second Language Research, 17*, 442–456.

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning*, *51*(Suppl. 1), 1–46.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275.

Ellis, R., Basturkmen, H., & Loewen, S. (2001a). Learner uptake in communicative ESL lessons. *Language Learning*, *51*, 281–318.

Ellis, R., Basturkmen, H., & Loewen, S. (2001b). Preemptive focus on form in the ESL Classroom. *TESOL Quarterly*, *35*, 407–432.

Ellis, R., Loewen, S., & Basturkmen, H. (1999). *Focussing on form in the classroom* (Occasional Papers, Number 13). Auckland, New Zealand: The University of Auckland, Institute of Language Teaching and Learning.

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

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Erlbaum.

Genesee, F. (1987). *Learning through two languages*. Boston: Heinle & Heinle.

Gravetter, F., & Wallnau, L. (1991). *Essentials of statistics for the behavioral sciences*. St. Paul, MN: West.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Greenhouse, J., Bromberg, J., & Fromm, D. (1995). An introduction to logistic regression with an application to the analysis of language recovery following a stroke. *Journal of Communication Disorders*, *28*, 229–246.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Rowley, MA: Newbury House.

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Hulstijn, J. H. (1995). Not all grammar rules are equal: Giving grammar instruction its proper place in foreign language teaching. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 359–386). Honolulu, HI: University of Hawai'i Press.

La Pierre, D. (1994). *Language output in a cooperative learning setting: Determining its effects on second language learning*. Unpublished master's thesis, University of Toronto (OISE), Toronto, Canada.

Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*, 33–51.

Lightbown, P. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 177–196). New York: Cambridge University Press.

Loewen, S. (2002). *The occurrence and effectiveness of incidental focus on form in meaning-focused ESL lessons*. Unpublished doctoral dissertation, The University of Auckland, Auckland, New Zealand.

Loewen, S. (2003). Variation in the frequency and characteristics of incidental focus on form. *Language Teaching Research*, *7*, 315–345.

Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Language Learning*, *54*, 153–187.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: Benjamins.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.

Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal, 82*, 357–371.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). New York: Cambridge University Press.

Lyster, R. (1998a). Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning, 48*, 183–218.

Lyster, R. (1998b). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*, 51–81.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*, 37–66.

Nabei, T., & Swain, M. (2001). Learner awareness of recasts in classroom interaction: A case study of an adult EFL participant's second language learning. *Language Awareness*, *11*, 43–63.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719–758.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition*, *17*, 459–481.

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pairwork. *Language Learning*, *50*, 119–151.

Pica, T. (2002). Subject-matter content: How does it assist the interactional and linguistic needs of classroom language learners? *Modern Language Journal*, *86*, 1–19.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction and research. In G. Crookes & S. M. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 9–34). Clevedon, UK: Multilingual Matters.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). New York: Cambridge University Press.

Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language Testing*, *9*, 173–186.

Saito, H. (1999). Dependence and interaction in frequency data analysis in SLA research. *Studies in Second Language Acquisition*, *21*, 453–475.

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press

Savignon, S. J. (2000). Communicative language teaching. In M. Byram (Ed.), *Routledge encyclopedia of language teaching and learning* (pp. 124–129). London: Routledge.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (Tech. Rep. No. 9, pp. 1–63). Honolulu, HI: University of Hawai'i Press.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
Slimani, A. (1989). The role of topicalization in classroom language learning. *System*, *17*, 223–234.
Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *30*, 73–87.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in it development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–252). Rowley, MA: Newbury House.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in the study of language* (pp. 125–144). Oxford: Oxford University Press.
Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.
Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, *18*, 275–302.
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. (1990). Attending to content and form in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, *12*, 287–301.
Williams, J. (2001). The effectiveness of spontaneous attention to form. *System*, *29*, 325–340.

APPENDIX

TRANSCRIPTION KEY

Symbol and meaning

L	Learner
T	Teacher
CAPITALS	Emphasis
(laugh)	Extra information
〈 〉	Inaudible
(2.0)	Timed pause
( )	Untimed pause
(.)	Micropause
=	Linked speech
?	Rising intonation
.	Falling intonation
-	Interrupted speech
. . .	Continuing discourse
:	Lengthening
[ ]	Phonetic representation
[	Overlapping speech
R	Researcher

https://doi.org/10.1017/S0272263105050163 Published online by Cambridge University Press