SSLA, **31**, 437–470. Printed in the United States of America.
doi:10.1017/S0272263109090366

TASK COMPLEXITY, FOCUS ON FORM, AND SECOND LANGUAGE DEVELOPMENT

Andrea Révész
Lancaster University

Tasks have received increased attention in SLA research for the past decade, as has the role of focus on form. However, few empirical studies have investigated the relationship among tasks, focus-on-form techniques, and second language (L2) learning outcomes. To help address this gap, the present study examined how the task variable +/– contextual support combined with the focus-on-form technique known as recasting affects L2 morphosyntactic development. The participants were 90 adult learners of English as a foreign language, randomly assigned to one of five groups: four comparison groups and a control group. The comparison groups differed as to (a) whether they received recasts while describing photos and (b) whether they could see the photos while describing them. The control group only participated in the testing sessions. A pretest-posttest-delayed posttest design was employed to detect any improvement in participants' ability to use the linguistic target, which was the past progressive form. Results from multifaceted Rasch measurement yielded two main findings. First, learners who received recasts but did not view photos outperformed learners who received recasts while viewing photos. Second, the group that viewed photos but did not receive recasts achieved greater L2 gains than the group who neither viewed photos nor received recasts.

I would like to thank ZhaoHong Han and James Purpura for their insightful comments on the research on which this article is based. I am also grateful to Alison Mackey, Monika Ekiert, Rebecca Sachs, Judit Kormos, and the five anonymous *SSLA* reviewers for their helpful suggestions on this article. Any errors, of course, are my own. This research was supported in part by the international research foundation for English language education (TIRF) and the Spencer foundation.

Address correspondence to: Andrea Révész, Department of Linguistics and English Language, Bowland College, Lancaster University, LA1 4YT, UK; e-mail: a.revesz@lancaster.ac.uk.

© 2009 Cambridge University Press 0272-2631/09 $15.00

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

There is now a widespread recognition among SLA researchers that (a) adult second language (L2) learning largely follows naturalistic learning processes (e.g., Long & Robinson, 1998; cf. DeKeyser, 1998) but (b) that grammar instruction that respects the learner's internal syllabus is useful and, at times, necessary if nativelike proficiency is the goal of L2 teaching (e.g., Long & Robinson; Sharwood Smith, 1986). Given these premises, the field of SLA research has seen an increased interest in task-based language teaching (TBLT) over the past decade (e.g., Bygate, Skehan, & Swain, 2001; Ellis, 2003). Unlike traditional teaching paradigms, task-based approaches seek to create an environment for universal acquisitional processes (i.e., implicit and incidental learning) to take place by promoting rich exposure to the target language (i.e., comprehensible input) and plentiful opportunities for meaningful communication (i.e., interaction and output). Moreover, these task-based approaches strive to combine a focus on meaning with a timely focus on linguistic forms—that is, shifts of attention to form consistent with the learner's internal syllabus.

Despite general agreement about the necessity of incorporating a focus on grammar into TBLT, L2 researchers' opinions seem to vary in terms of what form the pedagogical intervention should take. Among the proposals that have been made so far, two have received the most attention: (a) Long and colleagues' suggestion to incorporate focus on form in TBLT as a methodological principle (Long, 2000; Long & Crookes, 1992; Long & Robinson, 1998) and (b) Skehan and others' (Robinson, 2001, 2007; Skehan, 1998; Skehan & Foster, 2001) approach of proactively manipulating task complexity, "the (inherent) cognitive demands of tasks" (Robinson, 2001, p. 287), as a means to induce learner attention to form during task performance (Skehan & Foster). Although these two proposals, separately, have been the subject of much research in recent years, few studies exist that have investigated the two together in the context of TBLT.

The present study intends to fill this gap by examining how task complexity combined with focus on form affects SLA.[1] More specifically, the relationships between the task variable contextual support, the focus-on-form technique known as recasting, and the acquisition of the English past progressive form are explored. First, a brief review of the relevant literature will be provided and the results will then be described and discussed.

BACKGROUND

Focus on Form

As defined by Long (2000):

> focus on form refers to how attentional resources are allocated and involves briefly drawing students' attention to linguistic elements [...] in context as they arise incidentally in lessons whose overriding focus is

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

> on meaning or communication. The temporary shifts in focal attention are triggered by students' problems with comprehension or production. (p. 185)

This definition of focus on form entails two critical characteristics: (a) Attention to form occurs in discourse that is primarily meaning-based and (b) attention to form takes place incidentally in response to learners' linguistic needs. For Doughty and Williams (1998), however, the term *focus on form* has a broader scope: It also includes planned attention to form—that is, instruction directed at preselected linguistic items in the context of meaningful language use (Ellis, Basturkmen, & Loewen, 2001). This broader definition was adopted here.

Focus on form can be achieved in a variety of ways. Implicit focus-on-form techniques seek to direct learners' attention to form unobtrusively, without overt reference to rules and forms, minimizing the interruption of the pedagogic intervention on the processing of meaning. Conversely, explicit strategies direct learners' attention to form more obtrusively, often exploiting metalinguistic information, and, as a result, are more apt to intrude on the flow of communication. Whereas research to date suggests that explicit focus-on-form techniques are more efficient in triggering immediate restructuring of mental representations and improvement in behavior than implicit techniques (Norris & Ortega, 2000), L2 researchers are devoting increasing attention to implicit strategies. In particular, it has been argued that implicit strategies are more capable of engendering the acquisition of form-meaning mappings (Doughty, 2003).

Of all the ways implicit focus on form can be implemented, corrective feedback moves, such as recasts, clarification requests, comprehension checks, and repetitions, have been the most investigated, and researchers have identified several positive functions associated with these strategies. Most importantly, such feedback may help learners notice gaps between their interlanguage (IL) and the target language system. Noticing, in turn, may trigger destabilization and subsequent restructuring of the L2 grammar (Gass, 1997; Schmidt, 2001). Another frequently cited benefit is that implicit corrective feedback techniques may push learners to modify their output and that these techniques may thus assist learners in gaining control over or extending their linguistic knowledge (McDonough, 2005; Swain, 1995).

Recasts

Recasts, one type of implicit corrective strategy, have received particular attention from SLA researchers. Recasting involves the reformulation of a learner's utterance by altering one or more incorrect forms therein while

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

retaining its semantic content. Example (1), obtained from data collected for the current study, illustrates this technique.

(1) Learner: *And I saw a boy next to the bar. I think he was with his girlfriend. They* ***talking*** *to each other.*
Recast: *They* ***were talking*** *to each other.*

Therefore, a recast can not only provide a targetlike model but also potentially indicate to a learner that his or her utterance has been erroneous. In this way, the target language and the erroneous IL utterances are juxtaposed, which may induce the learner to notice the gap and, subsequently, to make a cognitive comparison between the targetlike form and his or her own incorrect utterance (Doughty, 2001; Ellis, 1995; Long, 2007). The availability of the model, some have argued, may even assist learners in acquiring new knowledge (e.g., Long). In other words, recasts may have the potential to facilitate the creation of initial mental representations of L2 forms. Another benefit associated with recasts is that they are reformulations of the learners' own utterances: Learners can therefore be expected to understand at least part of each recast given in response to their utterances because recasts are predicated on what they have just said. Thus, the learners will need to devote less attention to deciphering the meaning of the message and, in turn, will be able to allocate more processing resources to form and to form-function mapping (Long). Finally, recasts are the least intrusive of the implicit corrective strategies and, as a consequence, are the most likely to draw learners' peripheral attention to form without diverting their focal attention from meaning (Doughty, 2001). This is an important advantage, because directing learners' focal attention to form has been claimed to hinder form-function mapping (Doughty, 2003). In contrast, drawing learners' "roving attention to linguistic forms while the remainder of selective attention remains engaged in processing meaning" (Doughty, 2001, p. 251) has been argued to facilitate mapping.

The results of existing empirical studies suggest that recasts can indeed have a beneficial impact on SLA (Mackey & Goo, 2007), but the degree to which they do depends on several factors (Ellis & Sheen, 2006; Long, 2007). Some of these factors are learner-internal variables such as age (Mackey, Oliver, & Leeman, 2003), level of proficiency (Lin & Hedgcock, 1996), working memory (e.g., Mackey, Philp, Egi, Fujii, & Tatsumi, 2002), and developmental readiness (e.g., Long, Inagaki, & Ortega, 1998; Mackey & Philp, 1998; Philp, 2003). External factors discussed in the literature include the learning context (e.g., Ellis et al., 2001; Loewen, 2004; Lyster & Mori, 2006; Sheen, 2004), the type of linguistic feature targeted (e.g., Jeon, 2007; Long), and the manner of recasting (Doughty & Varela, 1998; Han, 2002; Loewen & Philp, 2006).

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Quite recently, task has also been identified as an important factor that alters the efficacy of recasts (Long, 2007; Robinson, 2001, 2003, 2005). For example, in a review article on recasts, Nicholas, Lightbown, and Spada (2001) noted that "the context or the task can influence the extent to which the recast is attended to" (p. 731). Similarly, Gass, Mackey, and Ross-Feldman (2005), in a study that investigated interactional patterns as a function of setting, indicated that interactional processes may differ according to the task. Moreover, in the most recent review of recasts to date, Long called for research to ascertain the robustness of the findings relating to recasts on different task types. Last but not least, Robinson (2001) also suggested that tasks, depending on their attentional demands, will differentially affect the efficacy of recasts.

Task Complexity and the Multiple-Resource Framework of Attention

In the current SLA literature, there are two competing accounts of how attention is deployed during task performance: the single-resource model (e.g., Skehan & Foster, 2001) and the multiple-resource model (e.g., Robinson, 2003; Wickens, 1992, 2007), each carrying a different set of implications as to what task variables may facilitate the noticing of morphosyntactic features. The attentional framework selected for this study was the multiple-resource view of attention. This model makes a distinction between separate resource pools along three dichotomous dimensions: processing stages (i.e., perception vs. response), modality (i.e., auditory perception with vocal response vs. visual perception with manual response), and codes of processing (i.e., verbal vs. spatial); each of these resource pools is responsible for a different aspect of task performance. The model (Robinson) holds that the relative ease or difficulty of a task will depend on the interference (i.e., confusion and cross talk) between similar codes or the competition for the same types of codes within specific resource pools along these three dimensions. In Wickens's (2007) words, interference will be higher "to the extent that any two tasks share common levels along more dimensions" (p. 187); that is, task performance is constrained by the potential breakdown in task-switching and scheduling operations due to interference and competition.

In addition to adopting a multiple-resource view of attention, Robinson (e.g., 2001, 2003) distinguished between two groups of task variables—resource-dispersing and resource-directing dimensions—to explain task effects. Increasing task complexity along resource-dispersing dimensions (e.g., by requiring learners to perform more than one task simultaneously

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

and without prior knowledge) disperses attention over many nonspecific areas of the L2. In contrast, increasing task complexity along resource-directing dimensions (e.g., by requiring learners to refer to events displaced in time and space) can direct learners' attention to specific, task-relevant features of the language code within the same resource pool. Tasks displaced in time, for example, will probably induce the use of past tense morphology. Robinson argued that increases in task complexity along resource-directing dimensions, as a result of a focus on specific forms, may induce learners to pay increased attention to input and output.

Task Complexity and Recasts

Robinson (e.g., 2001, 2005), within the scope of the cognition hypothesis of task-based language learning, proposed a series of predictions as to how increases in task complexity along resource-dispersing or resource-directing dimensions, or both, may affect L2 development and modulate the effectiveness of focus-on-form interventions. With specific reference to recasts, Robinson and Gilabert (2007) recently made explicit that this model predicts that increasing task demands "along resource-directing dimensions, and in general too along resource-dispersing dimensions" (p. 167), will lead to more extensive noticing, a greater amount of uptake, and longer term retention of input made salient by recasts. As Robinson (2001) explained, "increasingly complex tasks may prompt learners to look for more and more help in the input, attending to facilitative forms made salient by teacher intervention using [...] focus on form techniques" (p. 304).

To date, there is little empirical research to offer insights on the validity of this prediction. Only one study (Révész & Han, 2006), set out to explore the relationship between recasts and task complexity. This pilot study, in which all experimental groups received recasts, examined whether changes in task content familiarity and task type would influence the efficacy of recasts. Although results from ANOVAs showed significant main effects for both independent variables, participants showed greater use of recasts when performing less complex tasks, contrary to Robinson and Gilabert's (2007) prognosis. However, given a number of methodological weaknesses (for instance, the fact that no true control group was included in the design) these findings can only be suggestive. One of the major goals of the present study, therefore, was to further explore the joint effects of recasts and task complexity on L2 learning while circumventing the weaknesses of Révész and Han's study.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Task Complexity and L2 Learning

A second major goal of the current study was motivated by Skehan and Foster's (2001) proposal for integrating a focus on form into task-based instruction. Based on their information processing approach, Skehan and Foster claimed that learners' opportunities to focus on form should be maximized through attentional manipulation of task variables in the context of TBLT. For Skehan and Foster, tasks can be designed and implemented such that learners will place enhanced attention on form or, to be precise, on the two performance dimensions of form: accuracy (i.e., using the target language correctly) and complexity (i.e., using advanced and elaborate IL structures). (Note that, here, the term *complexity*, when used to describe linguistic performance, refers to syntactic as opposed to lexical complexity.) Balanced development in morphosyntax will thus be achieved by sequencing tasks in such a way that focus on one aspect of form (e.g., complexity) is reinforced by focus on the other (e.g., accuracy). In other words, Skehan and Foster contended that proactive manipulation of task design can promote L2 learning. Ellis (2005) also appeared to share this position, arguing that task variables that have the capacity to promote an increased focus on language form during performance are also likely to have the capacity to lead to greater IL development. Similarly, Robinson's (2001, 2007) cognition hypothesis suggests that task complexity manipulations may have direct effects on L2 learning.

In particular, Robinson (e.g., 2005) predicted that more complex tasks along resource-dispersing dimensions will facilitate automatization of, and real-time access to, an already established and developing IL system, as these tasks increasingly simulate real-world procedural and performance conditions. In contrast, more complex tasks along resource-directing dimensions and the increased conceptual and cognitive demands they entail will direct learners' attention to the overlaps and divergences of the conceptual systems of the first language (L1) and the L2 as well as to the ways in which concepts are grammaticized in the two languages, which will eventually lead to the development of new L2 form-function mappings. More specifically, increasing task complexity along resource-directing dimensions, among other things, will lead to heightened attention to speech production, which will result in greater accuracy and complexity of L2 output as well as more extensive IL analysis of task-relevant L2 constructions.

Whereas it appears plausible that manipulating the complexity of tasks along resource-directing dimensions has the potential to facilitate IL analysis, it is less obvious that more complex (resource-directing) tasks will always provide sufficient psycholinguistic conditions for analytical processes to be triggered. In terms of the multiple-resource

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

model, resource-directing dimensions rely on the same, verbal resource pool (Robinson, 2003; see also Kormos, 2000). Thus, when the increased conceptual demands posed by a task draw on form-function relations grammaticized differently in the L1 and the L2, for example, the competition and interference that may ensue between codes might need to take place in this resource pool.[2] In other words, enhanced complexity along resource-directing dimensions, although capable of directing learners' attention to specific areas of the L2, might at times, due to interference between L1 and L2 cues, simultaneously result in competition and interference in the verbal resource pool and, hence, inhibit these processes of IL analysis.

One might argue that in such cases—and in general—decreasing task complexity along resource-dispersing dimensions may help generate the sufficient psycholinguistic prerequisites for the creation of new form-function mappings. The decreased demands on task-switching and the scheduling mechanisms that the manipulation of tasks along such dimensions have the capacity to induce can free up attentional resources in the verbal resource pool, among others.[3] This, in turn, might enable learners to allocate more attention to aspects of the L1 and L2 systems needed to meet the conceptual and linguistic demands posed by a task and, therefore, increase the likelihood that the predicted positive effects of task complexity along resource-directing dimensions are achieved. Similarly, Robinson (2005) noted that the impact of task manipulations along resource-directing features "can be expected to be stronger when the task is simultaneously simpler along one or more resource-dispersing dimensions" (p. 7). So it would appear that, in addition to ensuring that a task is adequately complex along resource-directing dimensions to facilitate IL analysis, decreases in task complexity along resource-dispersing dimensions may also be facilitative of, and occasionally necessary to promote, processes linked to the acquisition of new form-function mappings.

As was the case for the link between recasts and task complexity, few empirical studies are available to confirm or refute the validity of this line of reasoning. Notably, however, the studies by Bygate (1996) and Gass, Mackey, Fernandez, and Alvarez-Torres (1999)—two empirical studies that have examined the effects of a resource-dispersing dimension task content familiarity (i.e., a form of prior knowledge) on short-term development of specific L2 features—found trends in line with this prediction. It is also worth noting that the most robust study that has investigated task complexity in relation to the learning of linguistic forms (Nuevo, 2006) did not detect an effect for task complexity; however, Nuevo focused on a resource-directing dimension. It is clear that more empirical studies are needed to elucidate the relationship between task variables and the acquisition of L2 forms. Another major goal of the present study, therefore, was to contribute to this underexplored area of research.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

+/– Contextual Support or the +/– Here-and-Now Dimension

This study explores the effects of manipulating task complexity along the +/– contextual support dimension, or the +/– here-and-now dimension, in relation to recasts and L2 morphosyntactic development. The –contextual support, or the there-and-then, condition was regarded as more complex, because tasks that require context-supported reference, such as narrating a story when picture prompts are available, are considered less cognitively demanding than those that require reference to objects or events without contextual support, such as narrating a story from memory, without picture prompts (Ellis, 2003; Robinson, 1995). The theoretical rationale for this premise is based on L1 and L2 acquisition findings that the ability to refer to events displaced in time and place emerges later than the ability to describe contiguous events (e.g., Bardovi-Harlig, 2000).

To use Robinson's (2001, 2005) terminology, the +/– here-and-now dimension of a task is resource-directing, because it has the capacity to draw learners' attention to specific L2 forms. For example, present time reference is more likely to be used under the here-and-now condition, whereas past time reference is more likely to be used under the there-and-then condition. As a result, the cognition hypothesis (Robinson, 2001, 2007) predicts that increased demands along this dimension will promote the efficacy of recasts as well as L2 learning.

Although there is little research that investigated the direct effects of the +/– contextual support dimension on the noticing of recasts and L2 development, several studies have investigated the effects of this dimension on oral production. In Robinson's (1995) study, participants were asked to narrate picture stories in the present tense while looking at picture prompts (here-and-now) and, also, to describe picture stories from memory in the past tense, after having seen the cartoon strips (there-and-then). Each participant performed three narratives altogether—at least one story with visual support and at least one without such support. Sequence (i.e., the order of presentation of the stories and the presence or absence of visual support), cartoon strip, and condition (+/– here-and-now) were counterbalanced. The results revealed an advantage for the there-and-then condition in terms of accuracy. These positive effects, however, did not prove to be significant in Robinson's study. A partial replication study by Rahimpour (1997), on the other hand, found this trend toward greater accuracy to be statistically significant. Similarly, Iwashita, Elder, and McNamara (2001), in a much larger scale study, observed that narrating picture stories under the there-and-then condition led to significantly greater accuracy than narrating picture stories under the here-and-now condition. Gilabert (2007) also found that there-and-then narratives resulted in significantly

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

more accurate language use than their here-and-now counterparts. Finally, Ishikawa (2007) detected greater accuracy on written there-and-then tasks. It is important to note that none of the studies that investigated oral performance revealed significant effects for syntactic complexity (cf. Ishikawa).

In contrast, a study by Skehan and Foster (1999) yielded an effect for linguistic complexity during oral production. In that study, learners in a watch-and-tell condition had to simultaneously watch and tell a Mr. Bean video (here-and-now and +contextual support), whereas learners in a watch-then-tell condition had to narrate the story after watching the video (there-and-then and –contextual support). Comparing the learners' performance under these two conditions, Skehan and Foster observed that the watch-then-tell condition led to greater complexity; however, there were no effects for accuracy.

In sum, in line with the cognition hypothesis (Robinson, 2001, 2007), these studies have demonstrated that there-and-then tasks can induce increased attention to form. Contrary to the cognition hypothesis, however, no study that examined oral production thus far has found significant effects for accuracy and complexity simultaneously. The positive impact of increased task complexity along the +/– here-and-now (resource-directing) dimension might have proved to be stronger in these studies had the tasks been concurrently made less complex along resource-dispersing dimensions. In fact, in the study by Gilabert (2007), albeit not reaching statistical significance, the positive effects of increased complexity along the +/– here-and-now dimension appeared to be enhanced when planning time was made available.

One explanation for the mixed findings may be that empirical research to date has typically treated the +/– here-and-now dimension and the +/– contextual support dimension as a monolithic entity. In terms of Robinson's (2001, 2005) task framework, however, these two variables appear separable and rather different in nature: The +/– here-and-now variable is argued to be resource-directing, whereas the variable contextual support seems to be resource-dispersing. Although the presence of contextual support (e.g., a photo) allows learners to focus their attention primarily on speech production, its absence requires an additional effort to memorize, and later retrieve, the details of the prompt employed in the task.[4] In other words, the two conditions appear to pose respectively single versus dual demands, a distinction that, in Robinson's task framework, is considered resource-dispersing.

As a consequence, the current practice of operationalizing the +/– here-and-now dimension as the availability or unavailability of contextual support during task performance appears problematic. When operationalized this way, the +/– here-and-now condition seems to constitute a composite task variable, distinguished not only by resource-directing characteristics but also by resource-dispersing characteristics.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Not less problematic would be if studies, when operationalizing the +/– contextual support dimension, would induce participants to narrate in the past tense when they have no contextual support and in the present tense when contextual support is available. Such an operationalization would make it impossible to determine whether it is the differential linguistic difficulty posed by the tasks or the presence versus absence of contextual support that is responsible for any effects observed. In the present study, to establish the impact of the variable +/– contextual support on its own, learners were prompted to use the same time frame under both the –contextual support and +contextual support conditions.[5]

RESEARCH QUESTIONS AND HYPOTHESES

The following research questions and hypotheses were addressed:

1. What are the combined effects of recasts and task complexity on L2 morphosyntactic development?

Based on the cognition hypothesis (Robinson, 2001, 2007), it was predicted that learners who had received recasts when contextual support was not available would show greater development than learners who had received recasts when contextual support was available.

2. Does task complexity, in the absence of recasts, have an impact on morphosyntactic development?

Again, based on the cognition hypothesis (Robinson, 2001, 2007), it was expected that learners who had performed tasks when contextual support was available would show greater development than learners who had performed tasks when no contextual support was available.

Note that *morphosyntactic development* was operationalized as improvement in the learners' ability to use the English past progressive form.

METHOD

Design

The study employed a pretest-posttest-delayed posttest design, with 90 participants randomly assigned to one of four experimental groups and a control group (see Table 1). The control group only participated in the pretest and posttests, whereas each experimental group took part in three treatment sessions between the pretest and the posttest. The

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Table 1. Design of the study

	Recast groups		Nonrecast groups		
Session	+Photo (n = 18)	–Photo (n = 18)	+Photo (n = 18)	–Photo (n = 18)	Control (n = 18)
Pretest	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –photo oral
Treatments 1, 2, 3	+Photo oral	–Photo oral	+Photo oral	–Photo oral	—
Posttest	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral
Delayed posttest	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral	Written +Photo oral –Photo oral

Note. "Written" corresponds to the written description task, "+photo oral" corresponds to the +photo oral description task, and "–photo oral" corresponds to the –photo oral description task.

experimental groups differed as to (a) whether they received recasts while describing photos during the treatment (recast groups vs. nonrecast groups) and (b) whether they had contextual support available while describing photos (+photo groups vs. –photo groups). *Contextual support* was operationalized as the availability versus unavailability of a photo during its description. At each testing session, three tasks were used to assess the extent of learning triggered by the respective treatments: a written description task, an oral description task with photo support, and an oral description task without photo support. The delayed posttest was only administered to half of the participants in each group. The rest of the participants completed an exit questionnaire after completing the posttest.[6]

The first research question was examined by assessing the performance of the +photo recast group against that of the –photo recast group, whereas the second research question was investigated by comparing the extent of development shown by the +photo and the –photo nonrecast group.

Target Form

Recasts in the current study targeted the past progressive form. The rationale for choosing this particular form was twofold. First, it is realized via a free morpheme (*was* or *were*) and a syllabic bound morpheme (*–ing*); thus,

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

it is physically salient (Goldschneider & DeKeyser, 2001). Second, it denotes grammatical tense and aspect, which means that it has some communicative value. It has been argued that recasts are more likely to draw learners' attention to linguistic forms that fit the criteria of being physically salient and meaning-bearing (Long, 2007).

The experimental tasks elicited only one context of the past progressive, generally considered to be the prototypical usage of this form: The past progressive referred to something in progress at a particular time in the past (e.g., *He was jogging at 7 o'clock last night*). As demonstrated by Bardovi-Harlig (2000), the past progressive is a developmental feature, which means that learners pass through relatively fixed developmental stages when learning it. Specifically, Bardovi-Harlig identified three stages, as illustrated in (2), in the emergence of this form:

(2) a. bare progressive (e.g., *walking*)
b. present progressive (e.g., *is walking*)
c. past progressive (e.g., *was walking*)

In light of Bardovi-Harlig's findings, participants were considered developmentally ready, and thus eligible for participating in this study, if the present progressive form had emerged in their IL. *Emergence* was operationalized as producing the present progressive form at least twice with unique lexical items in a past progressive context on the pretest. This is similar to the criterion used in other acquisition studies that considered emergence (e.g., Mackey, 1999; Philp, 2003). Participants who showed accurate use of the target form on the pretest were eliminated from the study.

Participants

The 90 participants were learners of English as a foreign language (EFL) enrolled in elementary or preintermediate language classes in three high schools in Hungary. The pedagogical approach adopted by the schools was a mix of focus-on-forms and communicative language instruction. The students were placed in their classes either based on the results of a placement test administered at the beginning of the school year or due to the successful completion of a prior-level course. The written part of the pretest was administered to 139 students from 11 intact classes. The classes were selected with the help of expert opinions of teachers from the three institutions. Only classes that had not received prior instruction on the past progressive were administered the pretest. Each participant had been in his or her current class for at least 6 months at the onset of the experiment. Of the initial 139 participants,

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

105 students, who showed no use of the past progressive form on the written pretest, took the oral pretest. Based on the overall pretest results, 95 students proved developmentally ready for the target form and thus eligible to participate; however, two students had to withdraw due to scheduling conflicts. Of the remaining 93 students, 90 randomly selected students were invited to continue the study.

For all five groups combined, there were 47 female and 43 male students. The participants' ages ranged from 15 to 20 (M = 16;87, SD = 1;42) and they were all native speakers of Hungarian. They had received between 0.5 and 8 years of English instruction prior to the study (M = 2.68, SD = 2.37). Most of the students had never visited an English-speaking country, with the exception of four students who had attended a 2-week English course in Great Britain. One-way ANOVAs run on the variable age and length of previous English study confirmed that there were no significant differences among the five groups with regard to these variables: $F(4, 85) = 0.381$, $p = .822$ and $F(4, 85) = 0.480$, $p = .751$, respectively.

Treatment

Tasks. The treatment tasks were contextualized in the hypothetical scenario that the participants were taking photos in a New York City neighborhood (e.g., Soho) exactly at a time when a crime (e.g., a bank robbery) happened in that area. During the experiment, the participants' task was to describe their photos to the researcher, who played the role of a police officer. The participants were told that they should describe each activity in the photos carefully, because the police would like to know what everybody was doing at the time the crime occurred. In each photo, at least three people were engaged in clearly identifiable activities, such as sitting, painting, or walking.

Three versions of the treatment task in two formats—one for the +photo groups and one for the –photo groups—were developed. The tasks were created using the computer program Microsoft PowerPoint. The first slide of each presentation displayed the title (e.g., *Soho*) for 5 s and was followed by a slide that contained the task instructions, which were visible for 2 min. Then, after the researcher had orally checked that the participants understood the task, 10 photos were presented. With each photo, a title appeared on the screen that indicated the time at which the photo had been taken. This title disappeared after 10 s under both the +photo and –photo conditions. Next, participants were asked to describe the photo in 40 s. The –photo groups could only see a blank screen while speaking, whereas for the +photo groups, the photo remained available on the screen.

This task was piloted with beginning-level L2 learners to determine the adequate time intervals allocated to view the photos. A 10-s initial

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

exposure to the photos was established to be appropriate in that it provided learners with just enough time to identify the activities but did not allow for engaging in other thought processes. This latter aspect was considered crucial because when more time was made available during the pilot study, learners under the +photo versus –photo conditions tended to be involved in distinct cognitive processes. As posttask interviews revealed, whereas learners in the –photo groups, for the most part, focused on memorizing the activities for later recall during the additional time, their counterparts in the +photo groups were primarily concerned with how they were going to formulate their subsequent speech. The rationale for allotting 40 s to describe each photo was that this time period, for most learners, proved sufficient to perform the task without feeling pressured.

The three versions of the task were piloted on 10 native speakers of English and 17 beginning-level Hungarian learners of English. The native baseline data indicated that the tasks generated approximately the same number of obligatory contexts for the target form and that the use of the structure was natural in each task. The three versions were also found comparable in terms of lexical variation (measured by type and token ratio) and syntactic complexity (measured by number of clauses per T-unit—that is, an independent clause with all its dependent clauses) based on both the native baseline and the EFL learner data.

Recasts. During the treatment sessions, the participants in the recast groups consistently received recasts from the researcher when they erroneously used the past progressive. The recasts were typically of the simple isolated declarative type; that is, they targeted a single error and were provided with falling intonation, without added emphasis on the targeted feature (Kim & Han, 2007; Lyster, 1998). A small number of recasts were also provided, albeit randomly, in response to other IL forms.

Assessment Tasks and Scoring

The pretest, the posttest, and the delayed posttest consisted of three tasks: a written picture description task, an oral photo description task with photo support, and an oral photo description task without photo support. Three different but comparable versions of each of the tests were developed. These were administered in a split-block design to avoid any potential task effects.

Written Picture Description Task. The goal of this task was to gauge the learners' ability to use the past progressive during a written production task. The inclusion of a written test in the design was considered important for two reasons. First, participants' performance on the

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

written task would reveal whether the effects of the oral treatment (or lack thereof) had transferred to the written modality. Second, it was deemed possible that some learners would produce more developed language in writing than they would orally (Bardovi-Harlig, 2000).

The picture the participants were asked to describe showed eight people engaged in various activities in a park. The written instruction contextualized the picture as having been taken at a particular time of the day and asked the participants to describe what the people were doing at that time. Three versions of the task were piloted on 6 native speakers and 15 EFL learners. The results demonstrated that each picture generated the same number of obligatory contexts for the target form, of approximately the same level of syntactic and lexical complexity. Ten minutes were allotted for the task.

The coding of the data consisted of four steps. The first step was to identify the obligatory contexts for the past progressive. Then it was determined whether any progressive marking had been produced in these contexts. Next, the data were analyzed in terms of four categories based on the developmental sequence for the progressive (Bardovi-Harlig, 2000). In obligatory contexts, participants received 3 points for using the past progressive, 2 points for using the present progressive, 1 point for the bare progressive, and 0 points for any nonprogressive form. The participants' total score was calculated using the formula provided in Equation 1:

$$T = \sum_{i=1}^{N} \mathrm{p}_i / N \times 3. \quad (1)$$

In Equation 1, T is the total score on the task, Σ is the sum, N is the total number of obligatory contexts, and p is the number of points received for each obligatory context. Thus, the total score demonstrated the mean proportion of targetlike and IL use of the past progressive. In addition to the total scores on each test, a pretest-posttest or a pretest-delayed posttest gain score was computed.

After all data had been coded and verified by a second researcher, intercoder agreement was determined by comparing the first rater's coding with that of the second rater. The second rater scored 20% percent of the data, randomly selected across the five groups. Cohen's kappa was .96, demonstrating strong intercoder agreement.

Oral Photo Description Tasks. The pretest, posttest, and delayed posttest each employed two oral production tasks: a photo description task with visual support and a photo description task without such support. These tasks were similar to the treatment tasks, except for being shorter in length: Instead of 10, participants were required to describe 5 photos. Six versions of the task in two formats—a +photo and a –photo version—were developed. The different versions were subjected to the same piloting procedures as the treatment tasks. The sequence of the six versions in

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

each group was counterbalanced within the testing sessions. Additionally, the order in which learners were asked to describe the photos under the +photo versus –photo conditions was counterbalanced.

The participants' oral production, altogether 87 h of oral data, was tape-recorded, transcribed, and then coded and scored following the same procedure as for the written picture description task. To ensure the reliability of the transcriptions and the coding, 10% of the data was randomly selected and also transcribed and coded by a second researcher. Intertranscriber agreement, calculated by dividing the total number of items transcribed identically by the total number of items transcribed, was high (.971). Cohen's kappa for intertranscriber agreement was also found to be high (.942). Intercoder agreement was determined by the same procedure used for the written description task. The Cohen's kappa values for the oral description task with versus without photo support were .94 and .90, respectively, indicating strong intercoder agreement.

Exit Questionnaire. The goal of the exit questionnaire was to obtain information concerning the participants' perspectives on their own experiences throughout the experiment. The questions required participants to write their responses in Hungarian. Out of the six questions asked, only one, provided in (3), is relevant to the present article.

(3) *Was it easier to describe the photos when you could or when you could not see them? Why?*

The analysis of the exit questionnaire data involved three phases. First, the researcher reviewed the learners' responses and identified emergent categories by annotating the data. Second, the resulting annotations were grouped into more general themes. Finally, the researcher double-checked all of the annotations and themes of the content analysis. The same three steps were repeated a month later to ensure the consistency of coding. The Cohen's kappa value demonstrated a high level of intracoder agreement (.92).

Data Collection Procedures

The data were collected over a time span of 5 months. Each participant attended six to seven sessions over the course of 6 weeks. On the first day of the experiment, the written description pretest was given to the participants during normally scheduled class times. As soon as the written pretest was completed, the researcher scored the tests and, based on the results, invited learners who appeared developmentally ready to acquire the target form to participate in the oral pretest. Learners

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

were subjected to the oral pretest individually on a different day within the same week. Based on the oral pretest data, students who proved eligible for the study were invited to continue. The written and oral pretest both lasted 10 min. The treatment, which consisted of three sessions, started approximately a week after the pretest and took place on three separate days over a 1-week period. The treatment sessions lasted 15 min each. The day after the last treatment session, the participants were administered the posttest, and half of the learners, randomly selected from each group, also received an exit questionnaire. The rest of the participants performed a delayed posttest 4 weeks later. Both posttests and the exit questionnaire were administered to participants individually. The posttests lasted approximately 20 min, and participants were given 30 min to complete the exit questionnaire. In an attempt to control for exposure to the target forms outside of the experiment, the teachers of the participating EFL classes were asked not to focus on the target form during the period of data collection.

Statistical Analyses

As a first step, descriptive statistics were calculated for each group's performance on the three assessment tasks of the pretest, posttest, and delayed posttest. Next, intercoder agreement was determined for the three testing tasks by comparing the two raters' coding and scoring of the data. To examine the effects of the treatment on the participants' ability to use the target form, many-facet Rasch measurement (MFRM) was employed (Linacre, 1989). MFRM is an extension of the simple Rasch model, which computes item and person estimates concurrently and produces measures for each of these estimates on a true interval scale, known as the *logit scale*. Unlike the simple Rasch model, MFRM allows not only item difficulty and person ability to be estimated but also the effects of other facets (i.e., definable aspects of the measurement condition) that may contribute to test score variation. MFRM, therefore, can be applied to analyze data obtained through experimental designs, among other things. Specifically, group can be specified as a facet in addition to participants and test items, and the impact, or the lack thereof, of the group facet on participants' pretest-posttest gains can thus be estimated on the same scale as that of the other facets.

In addition to calculating logit estimates for each facet, MFRM analysis also computes the significance of any differences that may exist among elements of a given facet (e.g., differences in gains among experimental groups). Another important feature of MFRM is that it provides fit statistics for each element, which indicate how well the data fit the stochastic expectation of the model. In the case of group

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

effects, for instance, the fit statistics show how consistently the facet group impacted participants' performance.

For the current study, the MFRM analysis was performed using Facets 3.61 for the IBM (Linacre, 2006). The model used for the analysis was the rating scale model, which assumes that the steps of a scale are equivalent across all elements of a given facet. The MFRM analysis was specified as having four facets: (a) participants' gains, (b) group, (c) time (posttest vs. delayed posttest), and (d) the difficulty of testing tasks. The mathematical model used for the analysis is provided in Equation 2:

$$\log(P_{njikx}/P_{njikx-1}) = B_n - C_j - D_i - E_k - F_{ix}, \quad (2)$$

where P_{njikx} is the probability of participant n achieving a gain score of x on testing task i in group j at time k, $P_{njikx-1}$ is the probability of participant n achieving a gain score of $x - 1$ on testing task i in group j at time k, B_n is the pretest-posttest gain of participant n, C_j = assignment to group j, D_i is the difficulty of testing task i, E_k = testing at time k, and F_{ix} = difficulty of achieving a gain score of x on a particular testing task i.

The analysis was conducted on the pretest-posttest and pretest-delayed posttest gain scores obtained from the written and oral description tasks. The data were originally modeled on a 100-point rating scale to correspond to the range of percentage scores obtained. However, the MFRM analysis yielded disordered average measures and step calibrations for this scale, which indicated problems with the functioning of the scale.[7] Therefore, the percentage scores were collapsed into a 6-point rating scale by recoding the data (Linacre, personal communication, November 26, 2006). Although rating scales with different step structures were also tested, the 6-point scale appeared best for representing the current dataset, because it included the most categories while still yielding ordered average measures and step calibrations.

The mean ability value of each group was anchored at 0 logits. In this way, it was ensured that there was enough connectivity in the dataset, which, due to the between-subjects design, would otherwise have contained disjoint subsets. Given that the participants were assigned to the experimental and control groups at random, the mean ability of the participants assigned to each group is assumed to be randomly equivalent. Except for the facet tasks, the mean difficulty of each facet was set at 0 logits.

RESULTS

Number of Recasts

As illustrated in Table 2, the mean number of recasts gradually decreased for both recast groups during the treatment, with the +photo recast group receiving a moderately greater number of recasts during each treatment

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Table 2. Number of recasts per treatment session by groups

Session	*M*	*SD*
+Photo recast group (*n* = 18)		
Session 1	14.89	7.78
Session 2	8.50	7.72
Session 3	6.39	6.84
–Photo recast group (*n* = 18)		
Session 1	11.44	5.36
Session 2	6.33	4.54
Session 3	3.06	2.80

session than the –photo recast group.[8] This indicates that both recast groups made fewer errors over the course of the treatment sessions and that the participants in the –photo recast group correctly produced the target form more frequently than their counterparts in the +photo recast group. In parallel, standard deviations for the +photo recast group were higher than for the –photo recast group, which demonstrate a larger degree of variability for the +photo recast group. Independent samples *t*-tests based on the number of recasts given to the +photo versus the –photo recast group during the treatment yielded no significant difference between the groups for any of the three treatment sessions: For the first session, $F(1, 34) = 2.39$, $p = .13$, for the second session, $F(1, 34) = 1.05$, $p = .31$, and for the third session, $F(1, 34) = 3.67$, $p = .07$.

Descriptive Statistics for the Dependent Variables

Oral Description Task With Photo Support. Table 3 presents the descriptive statistics for the scores on the oral description task with photo support. Both recast groups exhibited a considerable gain from the pretest to the posttest, with the –photo recast group demonstrating moderately larger gain than the +photo recast group. The +/– photo nonrecast groups showed a small increase of .08. The control group displayed no change from the pretest to the posttests. On the delayed posttest, the +photo recast group and the –photo nonrecast group maintained their respective gains. In contrast, the –photo recast group and the +photo nonrecast group showed respectively a slight decrease and increase in their use of the past progressive.

Oral Description Task Without Photo Support. The descriptive statistics for the scores on the oral description task without photo support are provided in Table 3. The –photo recast group exhibited the largest gain,

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Table 3. Descriptive statistics for the oral description task with and without photo support

	+Photo			–Photo		
Test	*M*	Mean gain[a]	*SD*	*M*	Mean gain	*SD*
+Photo recast group						
Pretest (*n* = 18)	.17	—	.10	.14	—	.13
Posttest (*n* = 18)	.80	.63	.17	.86	.72	.16
Delayed posttest (*n* = 9)	.79	–.01	.15	.86	.00	.14
–Photo recast group						
Pretest (*n* = 18)	.09	—	.09	.11	—	.09
Posttest (*n* = 18)	.93	.84	.11	.93	.82	.11
Delayed posttest (*n* = 9)	.84	–.09	.14	.88	–.05	.13
+Photo nonrecast group						
Pretest (*n* = 18)	.22	—	.11	.22	—	.12
Posttest (*n* = 18)	.30	.08	.17	.29	.07	.16
Delayed posttest (*n* = 9)	.39	.09	.23	.35	.06	.19
–Photo nonrecast group						
Pretest (*n* = 18)	.23	—	.14	.21	—	.13
Posttest (*n* = 18)	.31	.08	.16	.34	.13	.18
Delayed posttest (*n* = 9)	.34	.03	.22	.37	.03	.21
Control group						
Pretest (*n* = 18)	.22	—	.13	.22	—	.10
Posttest (*n* = 18)	.23	.01	.12	.25	.03	.08
Delayed posttest (*n* = 9)	.20	–.03	.11	.19	–.06	.12

Note. The maximum score was 1.0 point.
[a]Mean gain refers to pretest-posttest and posttest-delayed posttest gains in the table.

closely followed by the +photo recast group. The –photo nonrecast group showed a substantially lower gain but somewhat higher than the +photo nonrecast group. The control group displayed a slight increase from the pretest to the posttest. On the delayed posttest, the +photo recast group maintained its gain, whereas the –photo recast group showed a slight loss. The +photo and the –photo nonrecast groups, on the other hand, exhibited a small increase compared to the posttest. The control group did not maintain its gain on the delayed posttest.

Written Description Task. The descriptive statistics for the written description task are provided in Table 4. As on the oral measures, the –photo recast group improved the most, followed by the +photo recast group, the –photo nonrecast group, and the +photo nonrecast group. The control group also showed a small increase compared to the pretest. All groups, however, exhibited a decrease from the posttest to the delayed posttest, with the –photo nonrecast group demonstrating the greatest decrease and the –photo recast group demonstrating the least decrease.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

The Facets Analysis

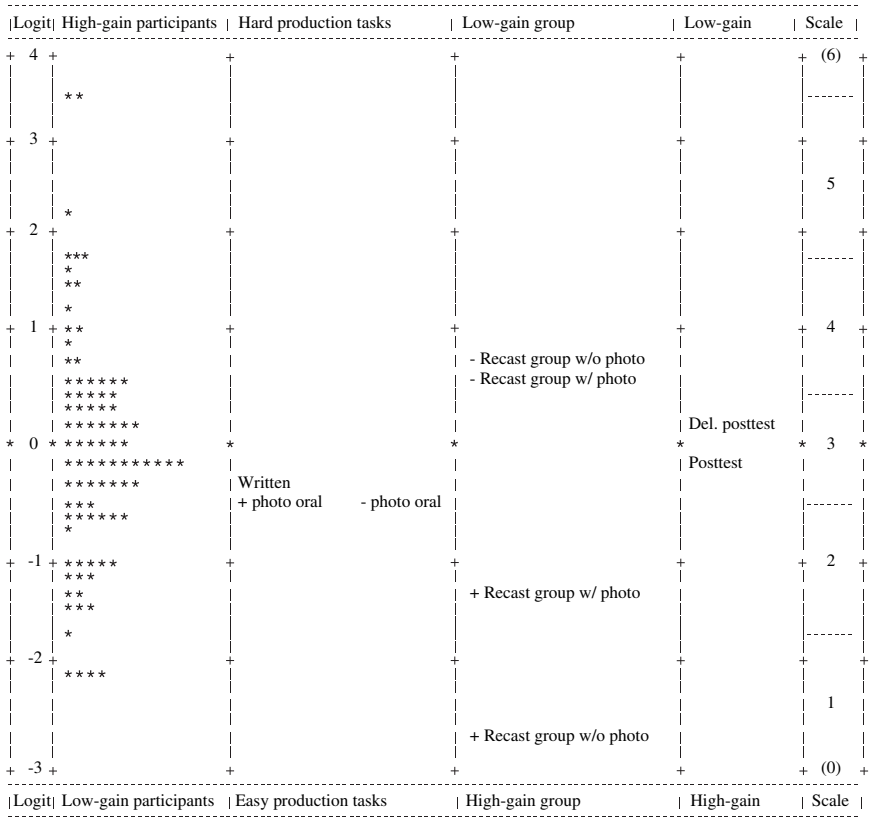
The summary map of the Rasch analysis is presented graphically in Figure 1. The first column in the figure displays the logit scale. The logit scale is an equal-interval scale, which provides a single frame of reference for all the facets of the MFRM analysis, which allows for comparisons both within and between the facets. The second column presents the distribution of the pretest-posttest gain scores on the written and oral description tasks in logits. Each asterisk (*) represents one participant's gain score. Participants with higher gains appear at the top of the column and participants with lower gains appear at the bottom of the column. The column shows that there was considerable variation in participants' gain estimates on the oral and written production tasks (approximately –2 to 3.5 logits). The third column represents the variation in the testing tasks in terms of difficulty. The tasks that appear higher in the column were more difficult to achieve high pretest-posttest gain scores on than the tasks that appear lower in the column.

Table 4. Descriptive statistics for the written description task

Test	*M*	Mean gain[a]	*SD*
+Photo recast group			
Pretest (*n* = 18)	.12	—	.18
Posttest (*n* = 18)	.92	.80	.18
Delayed posttest (*n* = 9)	.81	–.11	.32
–Photo recast group			
Pretest (*n* = 18)	.03	—	.04
Posttest (*n* = 18)	.97	.94	.08
Delayed posttest (*n* = 9)	.94	–.03	.15
+Photo nonrecast group			
Pretest (*n* = 18)	.08	—	.13
Posttest (*n* = 18)	.26	.18	.25
Delayed posttest (*n* = 9)	.17	–.09	.19
–Photo nonrecast group			
Pretest (*n* = 18)	.09	—	.13
Posttest (*n* = 18)	.35	.26	.24
Delayed posttest (*n* = 9)	.17	–.18	.16
Control group			
Pretest (*n* = 18)	.07	—	.11
Posttest (*n* = 18)	.17	.10	.18
Delayed posttest (*n* = 9)	.09	–.08	.14

Note. The maximum score was 1.0 point.
[a]Mean gain refers to pretest-posttest and posttest-delayed posttest gains in the table.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press



"Written" corresponds to the written description task, "oral" corresponds to the oral description task, "w/photo" corresponds to with photo support, and "w/o photo" corresponds to without photo support. The first column displays the logit scale. The second column presents the distribution of the pretest-posttest gain scores on the written and oral description tasks. Each asterisk (*) represents one participant's gain score. The third column represents the variation in difficulty of the testing tasks. The fourth column compares group gains. The fifth column compares participants' pretest-posttest gain scores with their pretest-delayed posttest gain scores. The sixth column graphically describes the rating scale.

Figure 1. Facets summary for the oral and written description tasks.

Thus, the third column indicates that the three production tasks were of approximately the same difficulty. The fourth column compares group gains. The groups that appear higher in the column displayed lower gains, whereas the groups that appear lower in the column exhibited higher gains. The fifth column compares participants' pretest-posttest gain scores with their pretest-delayed posttest gain scores. On the test that appears higher in the column, participants were more likely to achieve gain scores than on the test that appears lower in the

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

column. Finally, the sixth column graphically describes the rating scale used to score participants' performance on the tests.

The Rasch results for group and time—the facets directly relevant to the research questions posed—are examined in detail here. The summary statistics for these facets appear in Table 5. Turning to a description of the results for the group facet, the gain estimates for the groups spanned from –2.83 to 2.73 logits, which yields a logit spread of 5.56 logits ($SD = 2.22$). The overall difference between the group estimates was significant, $\chi^2(4, N = 90) = 452.00$, $p < .001$, with a separation reliability of .99. These statistics indicate that the five groups' pretest-posttest gains reliably differed from each other. The –photo recast group achieved the highest gain, followed by the +photo recast group, the +photo nonrecast group, and the –photo nonrecast group. As Table 5 indicates, although the difference between the recast groups' gains was relatively large, the nonrecast groups' gains differed only to a small degree. The analysis yielded the lowest gain estimate for the control group, which suggests that the experimental groups' pretest-posttest gains were, at least in part, a result of their respective treatments. As per the infit statistics (which indicate the extent to which the observations fit the modeled expectations, weighted to provide increased value to on-target observations), the infit mean-square mean was 1.83 ($SD = .46$) for the facet. Hence, following Pollitt and Hutchinson's (1987) criteria (i.e., two standard deviations away from the mean), any value outside the range of 0.91–2.75 would have been considered misfitting. All elements of the facet, however, had an infit value inside this range.

Table 5. Summary of statistics for the group and time facets ($N = 90$)

Facet	Measure logit	Model error	Infit mean square
Group			
+Photo recast	–1.64	0.16	2.51
–Photo recast	–2.83	0.20	1.86
+Photo nonrecast	0.78	0.13	1.83
–Photo nonrecast	0.96	0.13	1.72
Control	2.73	0.14	1.21
M	0.00 (2.22)	0.15 (.03)	1.83 (.46)
Time			
Posttest	–0.17	0.08	1.55
Delayed posttest	0.17	0.11	2.01
M	0.00 (.23)	0.10 (.02)	1.78 (.32)

Note. The *SD*s are in parentheses.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Moving on to the results for the time facet, the difficulty estimate for the delayed posttest was 0.17 logits, whereas the difficulty estimate for the posttest was –0.17 logits, with a standard deviation of 0.23 logits. Thus, the overall difficulty span between the posttest and delayed posttest measures was relatively small (.34). The reliability of the separation (.82) was moderately high, and the fixed chi-square test was significant, $\chi^2(1, N = 90) = 5.60$, $p < .05$. In other words, there was a significant difference between participants' performance on the posttest and the delayed posttest. Specifically, the participants showed higher gains on the posttest than on the delayed posttest. Neither element of the time facet was identified as misfitting or overfitting; the infit values for both tests were within the acceptable range (Pollitt & Hutchinson, 1987).

Exit Questionnaire

In response to the question of interest here, the large majority of participants (29 of 33) reported that they found it less difficult to describe the photos when they were able to view them. The rest of the participants (12.1%) felt that the availability or lack of contextual support did not make a difference in terms of task difficulty. It is important to note that not a single learner perceived the +photo condition as more demanding than the –photo condition. Most participants (72.4%) explained the lesser difficulty of the +photo condition by referring to the advantage of not having to remember or memorize what was in the photo. A subset of the participants also specified the benefit of not having to focus on memorization (51.5%). Most of these learners (42.1%) reported that when the photo was present, they were able to pay greater attention to the details of the photos and to describe all the activities portrayed, which they may not have been able to recall without having the photos in front of them. In other words, these learners felt that the presence of the photo freed up attentional resources, which, in turn, enabled them to complete the task more successfully (i.e., to describe the activities in the photos in as much detail as possible). Among the learners who specified the benefit for not having to remember the photo, a subset (9.4%) indicated that the increased attentional resources available under the +photo condition allowed them to focus on how to describe what they saw in English to a greater degree; that is, they were able to concentrate more on how to formulate their speech.

DISCUSSION AND CONCLUSIONS

The first hypothesis posited that learners who had received recasts in the absence of contextual support would show greater development in

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

their ability to use the target feature than learners who had received recasts when contextual support was available. The descriptive statistics showed a benefit for the –photo over the +photo recast treatment on each of the three testing tasks on both posttests. The MFRM analysis also found that receiving recasts in the absence of contextual support was more useful in terms of L2 morphosyntactic development than receiving recasts in the presence of contextual support. Therefore, these results uniformly confirm both the first hypothesis and Robinson and Gilabert's (2007) prediction concerning the link between task complexity and the effectiveness of recasts.

The cognition hypothesis (Robinson, 2001, 2007) provides one possible explanation for these findings. In terms of Robinson's (e.g., 2001, 2005, 2007) taxonomy for task features, the –photo condition is arguably more cognitively demanding in that it exerts dual versus simple task demands on learners. Therefore, the cognition hypothesis predicts that increasing task complexity along this dimension will disperse attentional resources and, thus, will make the learners' task more complex, which, as a result, will induce them to look for more and more external assistance during task performance. Robinson (2001) implied that focus-on-form techniques such as recasts, by directing learner attention to facilitative forms in the input, may be perceived by learners as providing such needed support. The exit questionnaire data confirmed that the –photo condition was perceived as more cognitively demanding than the +photo condition along the resource-dispersing dimension. The majority of the participants felt that describing the photos without contextual support was more difficult, because it forced them to simultaneously focus on speech production and memorization, which, in turn, made it more challenging to concentrate on task completion. In other words, as expected, learners perceived the –photo condition as more difficult than the +photo condition with respect to number of task demands (i.e., +/– single task feature). Hence, Robinson's proposal (see also Robinson & Gilabert, 2007), which states that increased task demands along the resource-dispersing dimension will result in greater intake and longer-term retention of input enhanced by recasts, appears compatible with both the quantitative findings and the questionnaire data.

A second potential explanation can be derived from Wickens's (2007) multiple-resource model. Given that, in terms of Wickens's model, the decoding of both the photos and the recasts required the learners to carry out perception tasks (i.e., processing visual and auditory input) as opposed to response tasks (e.g., speaking), it is possible that these two tasks entered into competition and, thus, attenuated each others' effects for the +photo recast group. In particular, the availability of the photo, although lessening memory demands, may have simultaneously acted as a distracter, drawing learners' attention away from the recasts. Following this line of thought, the absence of the photo in the –photo

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

recast group might have had contradictory effects on those learners' performance. On the one hand, because no contextual support was available for participants in this group, they did not need to handle the potential competition between the task of decoding recasts and that of decoding visual information, an advantage over the +photo recast group. On the other hand, the unavailability of the photo is likely to have dispersed the learners' attentional resources in another way, given that they concurrently had to focus on the recasts and recall the visual prompt. This, in turn, instead of inducing increased attention to input (Robinson, 2001), might have made it less likely for learners to notice recasts due to the increased demands on scheduling and task-switching operations. In this case, the superior pretest-posttest gains of the –photo recast group could be taken to suggest that the positive effects of recasts were greater in magnitude than those afforded by the presence of a photo. Such a difference in magnitude in fact was apparent in the results of the Rasch analysis; they indicated greater effects for recasts than for contextual support.

Assuming that there was in fact a competition between the recasts and the visual support under the +photo recast condition, a question arises: Why did the +photo recast group prioritize attending to the photo over attending to the recasts during task completion? One possible explanation lies in the monologic, as opposed to interactive, nature of the treatment task. Throughout the treatment, the participants' task was to describe the activities in the photos in as much detail as possible to the researcher. In other words, the task was essentially monologic in nature; the researcher was not assigned an active role during task performance. Therefore, there is a possibility that the learners perceived the external support offered by the photo as more relevant to task fulfillment than the external support offered by the recasts. The situation might have been different had the task required interaction between the participants and the researcher. Then the learners would, most likely, have been prompted to focus more on the recasts, as the interactive conditions afforded by the task would have necessitated attending to the interlocutor's speech and, hence, the feedback therein. If the present interpretation proved correct, the findings of this study would permit the speculation that task complexity may differentially affect the efficacy of recasts depending on whether the recasts are provided during monologic or interactive task performance. If further research were to confirm this speculation, the cognition hypothesis (Robinson, 2001, 2007) might need to be refined to accommodate the effects of interactional factors, such as monologic or interactive task conditions, in relation to the noticing of, and subsequent learning induced by, interactional feedback.

Finally, a third possible account of the findings regarding task complexity and recasts is discussed. Although both the +photo and –photo conditions were designed such that the past progressive form was

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

relevant to task completion, it could be argued that the learners perceived the use of past time reference as more natural in the absence of a photo, when they had no currently shared context with the interlocutor. As a consequence, they might have been more oriented toward using the past progressive form under the –photo condition, resulting in greater harmony between the focus of the recasts and the subject of task-induced attention. This interpretation, however, appears less plausible for two reasons. First, baseline data indicated that native speakers relied on past time reference rather than present time reference when describing the photos, regardless of whether visual support was available. Second, a post hoc analysis revealed that participants produced a relatively small percentage of verbs with present tense marking in past progressive contexts during the testing sessions (+photo: $M = 7.98\%$, $SD = 0.18$; –photo: $M = 8.13\%$, $SD = 0.18$) and treatment sessions (+photo: $M = 8.91\%$, $SD = 2.17$; –photo: $M = 14.83\%$, $SD = 2.63$), and such forms did not occur with significantly higher frequency under either of the testing or task conditions, as indicated by dependent and independent samples t tests for the testing sessions, $t(1,88) = -0.157$, $p = .876$, and the treatment sessions, $t(1,70) = -1.04$, $p = .168$, respectively. Furthermore, learners often relied on past time adverbials (e.g., *last Sunday*) while performing tasks with visual support, which suggests that they were in the past tense frame when talking about the photos.

The second hypothesis predicted that learners who had performed tasks with contextual support but without recasts would show greater development in their ability to use the target feature than learners who had performed tasks when neither contextual support nor recasts were available. The results indicated a weak confirmation of this hypothesis. Although the descriptive statistics, overall, showed slightly greater benefits for the –photo nonrecast treatment on the posttest, the +photo nonrecast treatment was found to be slightly more beneficial on the delayed posttest. The MFRM analyses conducted on the oral and written description tasks also confirmed a small advantage for the +photo nonrecast treatment; that is, the presence of contextual support, according to the majority of the measures, facilitated these learners' gains in their ability to use the past progressive form to a small degree.

As for the specific trends observed in relation to task complexity, the facets results for the production tasks provide support for the predictions derived from the cognition hypothesis (Robinson, 2001, 2007). As predicted, decreases in task complexity along the resource-dispersing dimension facilitated, even if to a small extent, the participants' development in the ability to use the target form. Again, given Robinson's (2003) assumption that decreases in task complexity along resource-dispersing dimensions can decrease the demands imposed on task-switching and scheduling mechanisms, it was predicted that

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

such task manipulations would promote development in the ability to use the past progressive form when tasks are sufficiently complex along resource-directing dimensions. This prediction was supported in the present study: Participants who performed tasks that were less complex along the resource-dispersing dimension showed slightly greater overall gains.

These results indicate that, as predicted by the cognition hypothesis (Robinson, 2001, 2007), task complexity can affect L2 learning outcomes. Nevertheless, the effects found for task complexity, again, were rather small, especially in comparison to the relatively large impact detected for the combined effects of recasts and task complexity. This discrepancy between the effect sizes may have been an artifact of the design used here. Arguably, the –photo and +photo conditions, as operationalized here, were not markedly different from each other. It is worth noting, however, that the study by Nuevo (2006), the most robust study that has previously attempted to investigate task complexity in relation to the learning of specific L2 features, yielded no significant effect at all for task complexity. Hence, it is also possible that task features, in and of themselves, without being combined with some type of external intervention, may only have a limited impact on L2 learning. One reason may be that whereas task-based production clearly has the capacity to promote greater control and automatized access to L2 forms, it is less likely to assist in the acquisition of new features or in triggering restructuring (Ellis, 2003). By implication, even if task complexity were to modulate the link between task-based production and L2 learning, it would only be expected to have a narrow effect, given the already limited potential of task-based production to lead to L2 development. Clearly, future studies are necessary to clarify the nature of the relationships among task complexity, production, and L2 learning.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

There are a number of limitations to this study that need to be acknowledged and considered in future research. First, a major weakness resides in the fact that recasts were treated as a constant. Although care was taken to provide recasts that were of the isolated declarative type and contained no added emphasis, factors such as length, single versus multiple corrections, and partial versus full reformulation were not held constant, all of which have been shown to have the capacity to modulate the effective functioning of oral feedback techniques (e.g., Loewen & Philp, 2006; Philp, 2003). A follow-up study could examine whether any of these factors did indeed have an impact on the efficacy of recasts.

Second, learner responses to recasts were not considered in the present analysis, and, given that modified output (e.g., Loewen,

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

2005; McDonough, 2005; Swain, 1995) as well as primed production (McDonough & Mackey, 2006) have been argued to positively affect L2 learning, a potential difference in the quantity and quality of the two recast groups' responses to feedback might have modulated the results. A follow-up study could assess whether modified output or primed production, or both, actually had an effect on the results obtained.

Third, the present research only investigated the effects of task complexity and its combination with recasts on a single linguistic feature—the past progressive form in English; thus, the results may not generalize to other structures or languages. A replication of this study with features of low, as opposed to high, physical salience or communicative value, or both, would be especially desirable, given that structures with such characteristics are the least likely to be noticed and, thus, learned solely from exposure to input (e.g., Long & Robinson, 1998).

Fourth, the treatment primarily elicited obligatory contexts for only one usage linked to the past progressive and supplied a limited number of contexts for other usages associated with this form. Hence, the treatment may have provided learners with a biased representation of the form-meaning mappings associated with the past progressive marking, which reinforced the general tendency of L2 learners to assume one-to-one correspondence between forms and meanings. A related limitation was that the assessments did not allow for testing how this restricted focus affected the acquisition of the additional meanings of the past progressive, given that the testing tasks mainly focused on a single meaning (i.e., the same usage of the target form).

Fifth, the exit questionnaire data, due to the inherent nature of the methodology, could only provide information on conscious processes. As a result, the exclusive use of this method to tap into the learners' perceptions limited the theoretical insights that could be gained regarding the cognition hypothesis, which puts heavy emphasis on constructs such as deep processing and automatic recognition processes in short-term memory—phenomena not available to conscious awareness. Future research, therefore, should circumvent this shortcoming by including implicit memory tasks in addition to protocols that tap explicit memory processes.

Follow-up studies that investigate the combined effects of recasts and task complexity as a function of individual difference variables are also warranted. As suggested by many, learner factors such as aptitude, motivation, and working memory, to name just a few, may significantly moderate the effects of task complexity and recasts on various aspects of L2 development (e.g., Mackey et al., 2002; Robinson, 2005; Trofimovich, Ammar, & Gatbonton, 2007). In the current study, differences in participants' working memory capacity may have played an important role given that the –photo condition required considerable memorization on the part of learners. Moreover, working memory has

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

been shown to be a significant predictor of learner noticing of recasts (Mackey et al.).

Another important avenue for future research would involve extending the research questions posed here to other tasks, conditions, and contexts. The present study included a single task type that involved learners in a monologic task in the context of individualized sessions. It is not necessarily the case that the effects found here would transfer to different task types, distinct interactive conditions, or classroom settings where learners typically receive less attention from the teacher.

Finally, in addition to continued efforts to investigate the interaction between task variables and the efficacy of recasts, future research should aim to explore the relationship between additional focus-on-form techniques and task complexity. This line of research could have important pedagogical implications: For instance, it might provide L2 educators with insight on how to design pedagogic tasks in order to enhance the effectiveness of focus-on-form interventions.

(*Received 17 September 2008*)

NOTES

1. Note that the terms *acquisition*, *development*, and *learning* are used interchangeably here.

2. Robinson (2003) acknowledged that such interference between L1 and L2 codes may be induced by task demands. In fact, using the competition model as a framework (e.g., MacWhinney, 2001), he implied that the presence and magnitude of such interference will depend on the extent to which the two languages differ in their relevant cues.

3. Of course, following Robinson (2003), it is also possible that decreasing task complexity along resource-dispersing dimensions will free up attentional resources in a different resource pool, but this is not necessarily the case. For example, allowing for planning time is, at least in part, likely to decrease attentional demands on the verbal resource pool.

4. With reference to storytelling tasks, Robinson (1995) also explained that whereas a context-supported condition only requires the speakers to describe episodes happening before their eyes, a context-unsupported condition involves the retrieval of stored events.

5. It might be argued that it is more natural to narrate in the present tense under the +contextual support condition. However, there is empirical evidence that showed that learners can be prompted to use past time reference when describing events even if contextual support is available (e.g., Han, 2002).

6. The rationale for administering either the delayed posttest or the exit questionnaire, but not both, to the participants was that completing the questionnaire prior to the delayed posttest might have raised learners' awareness of the purpose of this study. Alternatively, if the questionnaire had been performed after the delayed posttest, the long time interval between the treatment and the delayed posttest might have resulted in memory decay on the part of learners, threatening the validity of the questionnaire data.

7. *Average measures* are defined as the average of the logit estimates for all participants in the sample who produced a particular score, whereas *step calibrations* are the difficulties estimated for achieving a particular test score over another. Both measures are expected to increase monotonically as a variable increases in size; disordered average measures and step calibrations suggest that a certain rating scale does not adequately represent the data.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

8. An anonymous *SSLA* reviewer raised the issue of whether the +photo group received more recasts because they produced more language. This question would warrant an investigation in a follow-up analysis.

REFERENCES

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136–146). London: Heinemann.

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

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). New York: Cambridge University Press.

Doughty, C. J. (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. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Oxford: Blackwell.

Doughty, C. J., & Varela, E. (1998). Communicative focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). New York: Cambridge University Press.

Doughty, C. J., & Williams, J. (1998). Pedagogical choices in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197–261). New York: Cambridge University Press.

Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly*, *29*, 87–103.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. Amsterdam: Benjamins.

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

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, *28*, 575–600.

Gass, S. M. (1997). *Input, interaction and the development of second languages*. Mahwah, NJ: Erlbaum.

Gass, S. M., Mackey, A., Fernandez, M., & Alvarez-Torres, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*, 549–580.

Gass, S. M., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, *55*, 575–611.

Gilabert, R. (2007). The simultaneous manipulation of task complexity along the planning time and +/– here-and-now dimensions: Effects on L2 oral production. In M. del Pilar García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44–68). Clevedon, UK: Multilingual Matters.

Goldschneider, J., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, *36*, 543–572.

Ishikawa, T. (2007). The effect of manipulating task complexity along the (+/– here-and-now) dimension on L2 written narrative discourse. In M. del Pilar García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Clevedon, UK: Multilingual Matters.

Iwashita, N., Elder, C., & McNamara, T. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning*, *51*, 401–436.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Jeon, K. S. (2007). Interaction-driven L2 learning: Characterizing linguistic development. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 379–403). Oxford: Oxford University Press.

Kim, J. H., & Han, Z. (2007). Recasts in communicative EFL classes: Do teacher intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 269–297). Oxford: Oxford University Press.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, *50*, 343–384.

Lin, Y. H., & Hedgcock, J. (1996). Negative feedback incorporation among high-proficiency and low-proficiency Chinese-speaking learners of Spanish. *Language Learning*, *46*, 567–611.

Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA Press.

Linacre, J. M. (2006). Facets (Version 3.61) [Computer program]. Chicago: MESA Press.

Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Language Learning*, *54*, 153–188.

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, *27*, 361–386.

Loewen, S., & Philp, J. (2006). Recasts in the adult L2 classroom: Characteristics, explicitness and effectiveness. *Modern Language Journal*, *90*, 536–556.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179–192). Amsterdam: Benjamins.

Long, M. H. (2007). Recasts: The story so far. In M. H. Long (Ed.), *Problems in SLA* (pp. 75–116). Mahwah, NJ: Erlbaum.

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *24*, 27–56.

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

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

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

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, *28*, 321–341.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, *21*, 557–587.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–452). Oxford: Oxford University Press.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, *53*, 35–66.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, *82*, 338–356.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181–209). Amsterdam: Benjamins.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69–90). New York: Cambridge University Press.

McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, *27*, 79–103.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, *56*, 693–720.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719–758.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Nuevo, A. (2006). *Task complexity and interaction: L2 learning opportunities and interaction*. Unpublished doctoral dissertation, Georgetown University, Washington, DC.
Philp, J. (2003). Constraints on "noticing the gap": Non-native speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*, 99–126.
Pollitt, A., & Hutchinson, C. (1987). Calibrated graded assessment: Rasch partial credit analysis of performance in writing. *Language Testing*, *4*, 72–92.
Rahimpour, M. (1997). *Task condition, task complexity, and variation in oral L2 discourse*. Unpublished doctoral dissertation, University of Queensland, Brisbane, Australia.
Révész, A., & Han, Z. (2006). Task content familiarity, task type, and efficacy of recasts. *Language Awareness*, *3*, 160–179.
Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, *45*, 99–140.
Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). New York: Cambridge University Press.
Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.
Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, *43*, 1–32.
Robinson, P. (2007). Criteria for grading and sequencing pedagogic tasks. In M. del Pilar García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Clevedon, UK: Multilingual Matters.
Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics*, *45*, 161–176.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
Sharwood Smith, M. (1986). Comprehension versus acquisition: Two ways of processing input. *Applied Linguistics*, *7*, 239–256.
Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, *8*, 263–300.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
Skehan, P., & Foster, P. (1999). The influence of structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93–120.
Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). New York: Cambridge University Press.
Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.
Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171–195). Oxford: Oxford University Press.
Wickens, C. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins.
Wickens, C. (2007). Attention to the second language. *International Review of Applied Linguistics*, *45*, 177–191.

https://doi.org/10.1017/S0272263109090366 Published online by Cambridge University Press