

Effects of manipulating task complexity on self-repairs during L2 oral production*

ROGER GILABERT

Abstract

This paper analyses the effects of manipulating the cognitive complexity of L2 oral tasks on language production. It specifically focuses on self-repairs, which are taken as a measure of accuracy since they denote both attention to form and an attempt at being accurate. By means of a repeated measures design, 42 lower-intermediate students were asked to perform three different task types (a narrative, and instruction-giving task, and a decision-making task) for which two degrees of cognitive complexity were established. The narrative task was manipulated along +/- Here-and-Now, an instruction-giving task manipulated along +/- elements, and the decision-making task which is manipulated along +/- reasoning demands. Repeated measures ANOVAs are used for the calculation of differences between degrees of complexity and among task types. One-way ANOVA are used to detect potential differences between low-proficiency and high-proficiency participants. Results show an overall effect of Task Complexity on self-repairs behavior across task types, with different behaviors existing among the three task types. No differences are found between the self-repair behavior between low and high proficiency groups. Results are discussed in the light of theories of cognition and L2 performance (Robinson 2001a, 2001b, 2003, 2005, 2007), L1 and L2 language production models (Levelt 1989, 1993; Kormos 2000, 2006), and attention during L2 performance (Skehan 1998; Robinson, 2002).

1. Introduction

Speakers and learners around the world are often faced with having to communicate in a second or foreign language for different reasons, in different contexts, and under different conditions. There is general agreement that most communicative experience can be broken into discrete tasks, such as narrating a story, asking for and giving directions, or expressing an opinion when working out a problem. In the last few years, both from an interactive (see Mackey

and Gass 2007; and Long 2000 for reviews) and a cognitive perspective (see Ellis 2005; Robinson 2005, 2007; Skehan and Foster 2001 for reviews), SLA researchers have tried to shed light on the kind of interactive and cognitive processes that underlie the performance, whether written or oral, of communicative tasks and how their interactive and processing demands may lead to learning and acquisition. Both through semi-experimental and experimental research, they have looked into how L2 learners perform and acquire language when they carry out a variety of task types under different conditions of performance.

From an information-processing perspective, the goal of this paper is to analyze the effects of manipulating the cognitive complexity of L2 oral tasks on self-repair behavior during monologic production. To analyze how performance is affected by the different degrees of cognitive complexity that task design imposes on the learner's mind, three different task types are used: a narrative task, an instruction-giving task, and a decision-making task. This paper specifically focuses on how learners attend to the way they encode their speech during performance and, more specifically, on the rate and amount of errors they detect and repair in their speech. As will be argued later, self-repairs are here taken as a measure of accuracy since they denote both attention to form and an attempt at being accurate. To measure the effects of Task Complexity on narrative performance, task design is complexified by asking learners to narrate a story without the visual presence of a prompt and in the past tense. In the instruction-giving task, cognitive demands are made more complex by increasing the number of elements in a map and by having students navigate along different axes. Finally, the opinion-giving task is made more demanding by complexifying the relations between the different elements included in the task.

2. Task complexity and attention to form

2.1. Task complexity

As defined in Robinson (2001a, 2001b, 2003, 2007, this volume), the Cognition Hypothesis of adult task-based language learning, has advanced a series of assumptions and predictions of how increasing the cognitive complexity of language learning tasks may affect performance and potentially lead to interlanguage development. As far as performative arguments are concerned, gradually increasing the cognitive complexity of tasks along *resource-directing* variables has the potential to draw learners' attention to the way certain concepts are grammaticized in the L2. Following Talmy (2000), he suggests that as demands are made higher, learners may gear their attention towards the similarities and differences in the way closed-class items (e.g., prepositions) structure certain concepts (e.g., time or motion). Complementing this argument is

the fact that for such cognitive comparison of the overlap or divergence between L1 and L2 form-meaning mappings to take place, tasks must be kept simple along resource-dispersing variables, since otherwise attention may not be efficiently allocated to enable such comparison. Also, increasing communicative and cognitive demands forces learners to push production, stretch interlanguage, and destabilize fossilized forms. In Robinson's view (2003: 65), this can be achieved by "pedagogic interventions which manipulate the design characteristics of tasks, and the sequence in which they are presented to learners, so as to increase their functional and conceptual demands, so prompting learners from the use of elementary devices to more complex ones".

With regard to developmental arguments, Robinson (2003: 52) proposes that increasing the cognitive demands of language learning tasks can recapitulate or reproduce "the ontogenic course of conceptual development in childhood". In his view, then, increasing cognitive complexity along resource-directing dimensions such as the number of elements in a task (e.g., from the description of a space with few topological marks and relations to a space with many topological marks and relations among them), the degree of displaced past time reference (e.g., tasks in the Here-and-Now or the There-and-Then), or the reasoning demands (e.g., from narrative description of tasks with successive actions that do not require causal reasoning to those which do) leads to second language development. Robinson (2003: 52) assumes Slobin's (1985) idea that adults retain a scale of conceptual complexity acquired in childhood, and predicts that increasing the cognitive demands of tasks may provide learners "with optimal, ontogenetically natural, contexts for making form-function mappings necessary to L2 development". These parallels in adult-child language development, Robinson argues, can help operationalize sequencing decisions in task-based, process-based, and content-based syllabi.

2.2. Findings regarding attention to form

By applying a number of operationalizations to different task types, an array of accuracy measures, and with somewhat mixed results, different studies have tried to capture the effects of manipulating Task Complexity on how learners attend to form. With regards to the +/-Here-and-Now dimension, which has been the most heavily researched variable so far, Robinson (1995a) found higher accuracy on complex performance as measure by the target-like use of articles in a narrative manipulated along the degree of displaced, past time reference. In a partial replication of Robinson's study, Rahimpour's (1997) results displayed a significant effect for Task Complexity on accuracy as measured by error-free T-units but not when measured by the target-like use of articles. Iwashita et al. (2001) showed accuracy to be positively affected as measured by error-free T-units. Gilabert (2005, 2007) found that increasing complexity

along the +/- Here-and-Now variable had positive effects on accuracy as measured by the percentage of self-repairs and the ratio of repaired to unrepaired errors, with no significant effects showing when measured by the number of error-free T-units or the target-like use of articles. Ishikawa's (2007) results showed a trend for error-free T-units and error-free clauses and significant effects when measured by means of target-like use of articles. As for +/- elements studies, in a study which manipulated the number of landmarks in an oral map task that learners could refer to and their degree of familiarity with the area, Robinson (2001a) did not find significant effects of increasing Task Complexity on accuracy as measured by the number of error-free T-units. Kuiken, Mos, and Vedder (2005) and Kuiken and Vedder (2007) found accuracy to be higher under complex performance in written tasks which were manipulated along the number of requirements learners were asked to consider in the performance of the task. Regarding +/- reasoning demands studies, one study Niwa (2000) used narratives with storylines that required either a one-step connection between events or several steps. She reported a positive effect of Task Complexity on accuracy as mediated by individual differences. In narrative tasks which were manipulated along the structure of their storyline, Nuevo (2006) found no significant effects of Task Complexity on accuracy. Finally, it is often the case that in Task Complexity studies presence or absence of effects of complex performance on accuracy is explained in terms of this dimension potentially being in competition with the other dimensions of production, that is, fluency and syntactic and lexical complexity, the three dimensions which are said to compete for attention.

3. Language production and self-repair behavior

In order to understand how learners attend to form during language production, I will draw on a well-established model of language production by Levelt (1989, 1993). Briefly explained, the speech production system advanced by Levelt consists of a number of autonomous components which are responsible for different aspects of speech production. These components include: the conceptualizer, a component that is responsible for generating and monitoring messages; the formulator, in charge of giving grammatical and phonological shape to messages and which feeds on the lexicon; the articulator, which specializes in the motor execution of the message; an audition or acoustic-phonetic processor, which transforms the acoustic signal into phonetic representations; and the speech comprehension system, which permits the parsing or processing of both self-generated as well as other-generated messages. The conceptualizer is in charge of both generating messages and monitoring the whole process of production, an operation that requires conscious attention. In Levelt's account, speakers make use of their speech comprehension system to listen to

and process their own speech in exactly the same way they listen to and process the speech of others. More specifically, there are three monitor loops in the production system. One checks the preverbal plan against intentions. Another one checks the internal, articulatory plan against the overall plan, and a third one that monitors overt speech by means of the acoustic-phonetic processor. Learners monitor both the meaning and the well-formedness of their productions. When a problem is detected, several options are available, such as simply ignoring the problem, revising the preverbal message, or generating a new message. As will be detailed further below, this will depend on the nature of the problem as well as on other contextual factors.

An exhaustive taxonomy of problem-solving mechanisms related to perceived deficiencies in one's own production, more commonly known as self-repair, is the one advanced by Kormos (1999).¹ Her framework brings together theories about psycholinguistic processes of production, L2 research findings regarding the development of automaticity, as well as theories of consciousness, awareness, and noticing. Kormos adopts Levelt's perceptual loop theory of L1 monitoring. In agreement with van Hest (1996), Kormos suggests that perceptual loop theory can be used to explain monitoring in the L2 with no major qualitative changes. Nonetheless, there are at least two phenomena that the perceptual loop theory does not explain. Firstly, despite having sufficient underlying L2 knowledge, a lot of errors in L2 speech are not noticed by L2 speakers. Secondly, certain types of errors go undetected. In Kormos' view, differences in monitoring are explained by the fact that monitoring needs attentional control. One explanation of this is that because attentional capacity is limited, attention used to compensate for the non-automatic processes in the L2 (e.g., grammatical and phonological encoding) limits the amount of monitoring that takes place. This explanation assumes, then, that because a lot of attentional resources are being used to encode messages both grammatically and phonologically in the L2, fewer resources are available for self-monitoring which, as a consequence, is reduced. This limited-resources view of attention, however, is challenged by current multiple-resources accounts of attention (Wickens 1989, 1992). An additional explanation by Kormos is that attention to monitoring depends on individual differences, which display different working memory capacities among speakers (Harrington and Sawyer 1992). Thirdly, in Kormos' view, increasing task demands deviates attention from monitoring. Again, this assertion deserves closer scrutiny since there are opposing views about how task demands affect monitoring. Skehan (1998; cf. Skehan and Foster 2001; Foster and Skehan 1996; Skehan and Foster 1997), for example, suggests that as language tasks are made more demanding by, giving learners less time to plan them, they consume capacity which affects monitoring, with negative consequences for accuracy. Robinson (1995b, 2001a, 2001b, 2003, this volume) agrees with Skehan that increasing task demands

along planning time, as well as other *resource-dispersing* dimensions of tasks, deviates attention from monitoring. He argues, however, that increasing task demands along other dimensions, such as the number of elements in a task, the reasoning demands they impose on speakers, and their degree of displaced, past-time reference actually draws learners' attention to the way they encode messages with positive consequences for accuracy.

Following the previous classification of repairs (Kormos 1998; Levelt 1983; Van Hest 1986), Kormos distinguishes between different repairs (D-repairs), appropriateness repairs (A-repairs), and error repairs. The first type of repair is the consequence of errors with the conceptualization of the message, either because the information has not been organized properly or because it has been inadequately encoded in the preverbal plan. This basic classification taken from the work of Levelt (1983) was extended by Kormos (1998) to include inappropriate information repairs, message replacement repairs, and message abandonment repairs. Appropriateness repairs are meant to resolve ambiguity, achieve precision, and maintain coherence with previously used words. To these three she adds Brédart's (1991) pragmatic appropriacy repairs and repairs for 'good language'. Finally, error repairs are the result of wrong formulation, and may be used to correct an inappropriate syntactic structure, a lexical problem, faulty morphology, or a phonetic error. It is worth mentioning that the identification of the reasons behind different and appropriate repairs (i.e., the type of error they are correcting) require the use of retrospective protocol analysis for their identification. Conversely, overt error repairs are more easily identified and do not require introspection for their identification and classification.

Findings in the L2 self-repair literature (Kormos 2006: 131) have identified a number of facts that are now widely accepted. Firstly, there exist a number of factors which affect the decision to self-repair, such as the accuracy demanded by a specific communicative situation (i.e., an oral exam), the perception of the speaker regarding how seriously a specific error may prevent communication, and whether repairing an error may take a toll on the fluency of the utterance. Secondly, while in the L2 it is often the case that speakers prioritize the correction of content rather than form, certain types of L2 learners may devote a similar amount of attention to both content and form because their formal and grammar-oriented training in such L2. Thirdly, as learners develop their mastery of their L2, rather than increasing or reducing the global frequency of self-repairs they shift from simple error-repairs to more complex ones at the level of discourse (Kormos 2006: 133). In the fourth place, studies on self-repairs in the L2 have demonstrated that the "well-formedness" rule applies when errors are repaired, with studies reporting that over 80 % of repairs being well-formed.

3.1. *Self-repairs as a measure of accuracy*

In Gilabert (2005, 2007), three arguments were advanced in order to justify the calculation of self-repairs as a measure of accuracy: firstly, self-repairs, whether other-initiated or self-initiated (Schegloff et al. 1977), denote students' awareness of form and can be interpreted as learners' attempts at being accurate (Kormos 1999). Lyster and Ranta (1997: 57), for example, suggest repairs generated by learners as a result of corrective feedback lead them both to automatize the retrieval of target language knowledge they already have and to revise their hypotheses about the target language. Swain (1998: 66) has also hypothesized that noticing a hole in their own interlanguage may lead learners to notice the gap by directing their attention to relevant input. All these functions of self-repairs have been said to potentially lead to acquisition, and they have been pointed out in order to defend the benefits of certain types of corrective feedback. Self-initiated repairs, on the other hand, serve the same purposes as other-initiated repairs only that they are not the result of corrective feedback but rather are spontaneously generated by learners themselves or, in Levelt's terms (1989), they are the result of the speakers' monitoring of their own speech. These measures present accuracy 'in process' as learners try to correct and improve their own speech. Thirdly, at least one study has reported a higher proportion of self-repairs under certain planning conditions. Yuan and Ellis's (2003: 17), when analyzing the effect of on-line planning on learners' production, have reported a higher frequency of reformulations and self-corrections when they are given sufficient time 'during' (as opposed to 'before') performance. Concerning the subjects in their study, they suggest that on-line planners "engaged more fully in searching their linguistic repertoires and in monitoring their speech production". As will be further detailed below, in this study repairs are classified into two broad categories: on the one hand, error repairs related to lexis and faulty morphology and syntax grouped together;² and on the other hand, different and appropriateness repairs.

Motivated by both findings in the Task Complexity and the L2 self-repair literature, the main goal of this study is to measure the impact of Task Complexity on L2 learners self behavior. It will specifically tackle the issue of how and whether limited attentional resources affect monitoring. Secondly, with some exception in the planning time literature (Skehan and Foster 1996; Foster and Skehan 1997), most Task Complexity studies have focused on a single task type. This study will test whether the same learners behave differently when performing different task types. Additionally, this research will also test the claim that learners self-repair in similar amounts at different levels of proficiency.

3.2. Questions and hypotheses

Question 1. Do learners self-repair more and more frequently when tasks demands are increased along the degree of cognitive complexity?

Question 2. Does Task Complexity have a different impact on self-repair behavior on different task types?

Question 3. Do learners at different levels of proficiency display different self-repair behaviors?

Hypothesis 1. All task types performed under complex conditions will trigger fewer errors and a significantly higher rate and number of self-repairs.

This hypothesis will be tested by calculating the number of errors per AS-Unit, the ratio of errors to words, the number of error-repairs per AS-unit, the ratio of error-repairs to words, the number of error and non-error repairs per AS-unit, the ratio of error and non-error repairs to words, the percentage of self-repairs, the ratio of repaired to unrepaired errors, and the corrected ratio of repaired to unrepaired errors.

Hypothesis 2. Both the narrative task and the map task will trigger a higher level of self-repairs than the decision-making task.

There is no directional hypothesis for the hypothesis advanced here. This prediction is based on the Cognition Hypothesis (Robinson 2001b, 2003, 2007) which suggests that attention will be drawn to form during complex performance as long tasks as are kept simple along resource-dispersing dimensions. It is believed that while the structured storyline of the narrative clearly indicates what needs to be communicated, as does the route marked on the map, in the decision-making task the communicate intention is to be created from scratch. This will be done by comparing the results of the nine measures on the different task types under both simple and complex performance.

Hypothesis 3. There will be no significant differences between groups at different levels of proficiency for any of the tasks.

This hypothesis is motivated by previous findings in the L2 self-repair literature (Kormos 2006) which suggest that global self-repair behavior is not affected by different levels of proficiency. This will be analyzed by comparing two groups statistically organized into different levels of proficiency.

4. Experimental design

A repeated-measures design was used in which the within-learner factor was Task Complexity. Repeated measures analyses of variance (ANOVA) of the nine dependent variables were carried out to measure the effects of Task Complexity and the differences among task types. One-way analyses of variance (ANOVA) are used to detect difference between different levels of proficiency. Given that in this experiment learners were asked to carry out six different tasks, it was predicted that practice or carryover effects might take place from

one task to another. Besides task type, the sequence of conditions under which the tasks were performed was also thought to potentially affect performance. Students were randomly assigned to one of the six groups in a Latin Square design. Affective perception was tested by means of an affective variable questionnaire which asked learners to rate difficulty, stress, confidence, interest, and motivation on a 9-point Likert scale as the one used by Robinson (2001a).

4.1. *Participants*

42 volunteers from two different institutions participated in the study. On the one hand, students coming from the Blanquerna Communication Studies program at Ramon Llull University, in Barcelona, Spain, had been placed in a lower-intermediate level by means of entrance level test. Students coming from the University of Barcelona were expected to have a higher level of proficiency since their degree is English language and literature. Students in both institutions had received instruction in English for approximately the same number of years. Learners' ages ranged between 18 and 40. 17 students from Ramon Llull University and 25 from Universitat de Barcelona participated in the study. X-lex and Y-lex placement tests measuring vocabulary size were used to establish their level of proficiency. These tests have been developed and widely tested by Meara and Milton (2003) at the University of Wales at Swansea. A one-way ANOVA displayed significant differences ($F = 29,6617, p = .000$) in proficiency between the two institutions. Additionally, discriminant K-means tests were used to assign each student to either a low proficiency or a high proficiency group regardless of the institution they came from. Discriminant K-means analysis assigned 13 students to the high proficiency group and 29 to the low proficiency group.

4.2. *Materials*

Narrative: Two comic strips, previously tested in Gilabert (2005), were used to elicit narrative discourse. They were thought to be especially useful for data collection because they were wordless comic strips, they contained a small number of characters who were involved in the action, and they had a clear climax and resolution. Regarding the +/- Here-and-Now distinction, this research followed Robinson's (1995a) operationalization. For Here-and-Now, learners were asked to narrate the story in the present while they looked at the strips. For There-and-Then, learners were asked to narrate the story in the past tense, and they were not allowed to look at the pictures as they performed the task.

*Instruction-giving task:*³ two city maps were used for the instruction-giving task. The simple version contained few easily distinguishable landmarks. Learners were instructed to leave a message in a friend's voice mail giving him or her instructions as to how to get to a news stand where their friend was meant

to buy a newspaper. The other instructions included posting a letter in the post office, and buying a bunch of flowers in a flower shop before going back to a subway station. To do so, they had to ask their friend to basically move along a single lateral axis (i.e., left, right, straight). In the complex version, learners had to instruct their friend to pick up their dog from the vet's, to buy food from a department store, and to exchange money in the bank before going back to the subway station. This time task design was manipulated to include many points of reference which also looked very similar to one another. They had to ask their friend to navigate along the lateral (i.e., left, right, straight on a number of streets), vertical (i.e., up and down stairs inside two buildings), and sagittal axes (i.e., from the front to the back of the store, which forced learners to take a more complex perspective). This operationalization was based on Cornell et al. (1994) and Chown, Kaplan and Kortenkamp (1995) and their constructs regarding landmark identification, path selection, direction selection, and abstract environmental overviews. It is important to note that to keep both the simple and the complex task simple at the level of resource-dispersing dimensions, the route their friend was meant to follow was marked on the map.

Decision-making task: For the decision-making task, a simplified 'Fire chief' task used in cognitive psychology was used. In this task, learners were presented with a building where a fire has broken out and where a number of people need to be rescued. As Evans and Marciniak (1987) suggest, complexity is related to the intricacy of systems. The higher the number of factors and relations within a system, and the higher the amount of element interactivity, the more complex this system is. Following Quesada, Kintsch and Gómez-Millán (2005) Complexity was increased so that most variables were not related to each other in a one-to-one manner. The problem in the complex version required from learners not just one decision, but a long series, in which early decisions condition later ones. In both versions of the task learners were instructed to specify the actions they would take, determine the sequence of their actions, and justify their choice for actions and their specific sequence. In the simple version, there were similar types of people (i.e., people with no particular roles) in the building who were faced with similar degrees of danger. Learners had plenty of resources (i.e., three fire trucks and a helicopter which could be used simultaneously) and few unconnected factors to deal with (i.e., the people in relatively safe places, the fire being relatively static, the smoke blowing away from the building). In the complex version, learners had to deal with specific types of people (e.g., a pregnant woman, an elderly man, an injured person, a hero) which it was thought would force them to refer to psychological states terms which in L1 acquisition emerge in the order physiological > emotional > desire > cognitive state terms (Lee and Rescorla 2002; Shatz, Wellman and Silber 1983). The factors in the task were also intricately related and dynamic (e.g., the different fires moving towards the people and the smoke

blowing into the building through the ventilation system) and they had fewer resources (i.e., a single fire truck), which was thought would force them to prioritize, and later justify, their actions.

4.3. *Procedures*

Data collection took place in a single one-hour session. Learners sat facing the researcher and, after a few minutes of small talk to achieve rapport and collect personal information, they were asked to carry out the tasks under the different conditions. Since the goal of this study was to measure Task Complexity independently from other variables, a very short pre-task planning time was given to them (up to one minute before starting each task). The researcher provided learners with one or two words (e.g., 'checkout girl' or 'straight ahead') which had been found to be problematic during piloting. The steps taken during the session include a brief explanation of the research, small talk to establish rapport, and personal information retrieval. Each set of two tasks, in the simple and complex version, was followed by an affective variables questionnaire (Example of Group A: simple narrative > complex narrative > questionnaire > simple decision-making > complex decision-making > questionnaire > simple instruction-giving > complex instruction-giving > questionnaire).

4.3.1. *Measures.* In order to establish percentages and ratios the number of error-repairs and the number of non-error repairs (i.e., different and appropriateness) were calculated against the number of errors,⁴ the number of words, and the number of analysis of speech units (AS-units). As for errors, errors with faulty lexis, morphology, or syntax were counted. Lexical and morpho-syntactic self-repairs were counted and classified as error-repairs and all other repairs (i.e., different and appropriateness repairs not springing from an error but from a change in the speech plan) were counted and classified as non-error repairs. Finally, AS-units, as defined by Foster, Tonkyn and Wigglesworth (2000) were also computed.

Measures were calculated as follows: the number of errors per AS-unit was obtained by dividing the number of AS-Units by the number of error-repairs; the ratio of errors to words was calculated by dividing the number of errors by the total number of words; the number of error repairs per AS-units was calculated by dividing the number of AS-units by the number of self-repairs; the ratio of error-repairs to words was calculated by dividing the number of error-repairs by the total number of words; the total number of repairs was measured by dividing the number of AS-units by the number of error- and non-error repairs taken together; the ratio of error- and non-error repairs to words was measured by dividing the total number of error- and non-error repairs taken together by the total number of words; the percentage of self-repairs was com-

puted by taking the number of self-repairs⁵ and dividing it by the total number of errors and multiplying the results by 100; the ratio of repaired to unrepaired errors was measured by dividing the number of repaired errors by the number of unrepaired errors and multiplying the results by 100; the corrected ratio of repair was used to compensate for differences in text length which may trigger very large differences in the numbers of errors, and was calculated by dividing the number of error self-repairs by the square root of twice the number of errors ($E\text{-repairs}/\sqrt{\text{Error} \times 2}$). In my view, the number of error-repairs per AS-unit, the ratio of error-repairs to words, the number of error and non-error repairs per AS-unit, and the ratio of error- and non-error repairs to words capture the frequency with which learners repair. The last three measures, on the other hand, indicate the amount of self-repairs that learners produce.

4.3.2. *Statistical instruments, analyses, transcription and coding.* Three different kinds of statistical analyses are used in this study: descriptive statistics, which provide information about means and standard deviations; repeated-measures analyses of variance (ANOVA) are used for the calculation of main effects, and pairwise comparisons to identify the exact location of differences; one-way ANOVAs are applied to analyze potential differences between learners at different levels of proficiency. Normality of the data was achieved by detecting outliers for each variable by means of box plots and eliminating them from the count. Significance levels were set at $\alpha = .05$. Following an anonymous reviewer's suggestion, α was reset at .001 by applying the Bonferroni procedure.⁶ Result tables for the three hypotheses report on both α .

The CA mode of CHILDES (MacWhinney 1995) was used for the transcription of the 252 tasks. This software allows for the automatic calculation of items (e.g., words or tags) in a text. Both intrarater and interrater measures were used in the transcription and coding of the narratives. The transcription of the narratives was carried out by the researcher and two research assistants. Intrarater reliability reached 97%, and interrater⁷ agreement out of a randomly selected sample of 10% percent of the data reached 91.5%.

5. Results

5.1. *Affective perception results*

As seen in Table 1, repeated measures ANOVAs for the five affective variables display a significant main effect for the perception of difficulty, stress, and confidence, whereas no significant main effects were found for interest or motivation. Although learners graded the three complex versions of the tasks as more difficult, pairwise comparisons showed that learners perceived the complex version of tasks to be significantly more difficult only for the instruction-giving map task ($p < .05$) and the decision-making ($p < .001$), but not for the

Table 1. Main effects for affective variables by task: difficulty, stress, confidence, interest, and motivation

Dependent variable	Df	Sum of squares	F-value	p-value	η^2
Difficulty	5	25.407	8.574	.000 ^b	.192
Stress	5	12.149	5.643	.000 ^b	.136
Confidence	5	11.397	11.397	.000 ^b	.165
Interest	5	3.343	1.029	.402	.028
Motivation	5	2.279	1.869	.102	.033

Df = Degrees of freedom; η^2 = partial eta squared (effect size); ^a = $p < .05$; ^b = $p < .001$

narrative task, which did not show any significant differences. Pairwise comparisons for stress showed differences to exist between the different task types under complex performance but not between the simple and complex version of any of the task types. Finally, results show that learners felt significantly less confident when performing the complex fire chief task, with no differences existing between simple and complex versions of the narrative or the map task. The fact that the more complex narrative task is not perceived as more difficult, as opposed to the findings of Gilabert (2005, 2007), may be explained by the fact that more advanced students participated in the study and these may have not perceived the task in the 'there-and-then' as more difficult. A one-factor ANOVA with proficiency as between-groups factor showed this to be the case, since it showed that the high proficiency group perceived the complex narrative as significantly ($p < .05$) easier than the low proficiency group. Two conclusions can be drawn from the perception questionnaire results: first, that the overall operationalization of tasks into simple and complex versions was confirmed by the perception of students; second, that the narrative task was perceived as significantly easier than the map and fire chief tasks, with no significant differences existing between these two.

5.2. Results of Hypothesis 1: Self-repair behavior as affected by task complexity

Only the percentage of self-repairs showed the interaction between sequence of presentation and Task Complexity to be significant ($p < .05$). Pairwise comparisons showed Group B to significantly repair more than all other groups. This effect, however, was not found with any of the other variables. It is still believed that significant differences in performance are due to the different degrees of Task Complexity and not to the sequence in which they were presented. As seen in Table 3, main effects were significant for all measures.

Table 2. Descriptive statistics for all measures under simple and complex performance

Dependent Variable	Narrative task				Instruction-giving map task				Decision-making fire chief task			
	Simple		Complex		Simple		Complex		Simple		Complex	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Errors per AS-unit	1.49	.83	1.24	.96	1.35	.77	.78	.40	1.24	.69	1.24	.91
Ratio errors/words	.10	.06	.06	.04	.08	.04	.06	.03	.08	.05	.08	.05
Error-repairs per AS-unit	.21	.14	.26	.20	.06	.07	.08	.06	.10	.10	.15	.13
Ratio error-repairs/words	.01	.01	.01	.01	.19	.13	.23	.11	.00	.00	.01	.00
All repairs per AS-unit	.35	.22	.43	.29	.00	.00	.00	.00	.33	.18	.32	.19
Ratio all repairs/words	.02	.01	.03	.02	.01	.01	.02	.00	.02	.01	.01	.00
Percentage of self-repairs	16.91	14.32	21.07	15.06	5.85	7.00	9.72	5.79	9.12	9.21	11.39	9.42
Ratio of repaired to unrepaired errors	21.85	22.44	34.05	30.47	6.16	7.85	12.07	8.98	10.37	11.41	13.14	11.42
Corrected ratio repaired/unrepaired errors	0.40	.34	0.54	.38	.22	.26	.39	.26	.22	.21	.38	.33

Hypothesis 1 was partially confirmed with different results being obtained for each task. As shown in Table 4, results for the narrative task suggest that, when performed along displaced past time reference, learners made significantly fewer mistakes as shown by the ratio of errors to words, but no differences were found when measured by the number of errors per AS-units. As for the rate of repairs, none of the measures displayed any significant differences between the two conditions. In terms of quantity of self-repairs, the ratio of repaired to unrepaired errors and its corrected version displayed significant differences between conditions, with students repairing more times in the complex version. Hence, for the narrative the hypothesis was only partially confirmed. As for the map task, making the task more complex by increasing the number of elements and the number of axes along which learners needed to navigate significantly affected the rate of repair and the number of repairs they carried out. This is shown by all measures except for the error-repairs per AS-unit. It can therefore be concluded that Hypothesis 1 was largely confirmed for the map task. With regard to the decision-making fire chief task, increasing reasoning demands by complexifying the relations among the elements in the task did not seem to affect either the number of errors, the rate or the amount of self-repairs. Significant differences between the two conditions were only found when calculating the ratio of repaired to unrepaired errors. Surprisingly, when error-repairs and non-error repairs were taken together, the results run against the prediction, with learners self-repairing more often in the simple version. Thus, it can be concluded that Hypothesis 1 was not confirmed for the decision-making fire chief task.

5.3. Results of Hypothesis 2: Performance among task types

In general, very similar patterns of behavior can be observed under simple and complex performance. When performing simple tasks, learners made more errors, repaired more frequently and a larger proportion of errors in the narrative task than in the map task or in the fire chief task. If we compare the narrative to the map task, all measures except for the number of errors per AS-unit show that learners repaired more frequently and repaired a larger amount of errors in the narrative than when performing the map task. If the narrative is compared to the fire chief task, a similar pattern of behavior can be found, with the narrative showing fewer errors and a higher proportion of self-repairs. Only two of the measures, the total number of error- and non-error repairs per AS-Units and the ratio of repaired to unrepaired errors do not show significant differences. As far as the comparison of the map task and the fire chief task is concerned, while under simple performance learners made similar numbers of errors and self-repaired in similar amounts, under complex performance they made less errors and self-repaired more often in the map task, with no differences existing

Table 3. Repeated measures ANOVAs for all measures

Dependent variable	Df	Sum of squares	F-value	p-value	η^2
Errors per AS-unit	5	2.137	4.608	.001 ^b	.123
Ratio errors/words	5	.005	7.824	.000 ^b	.196
Error-repairs per AS-unit	5	.866	14.241	.000 ^b	.308
Ratio error-repairs/words	5	.001	16.653	.000 ^b	.357
All repairs per AS-unit	5	.286	10.036	.000 ^b	.228
All repairs per 100 words	5	.002	13.306	.000 ^b	.281
Percentage of self-repairs	5	701.276	9.059	.000 ^b	.258
Ratio of repaired to unrepaired errors	5	1692.499	7.675	.000 ^b	.242
Corrected ratio of repaired/unrepaired errors	5	.480	7.094	.000 ^b	.208

Df= Degrees of freedom; η^2 = partial eta squared (effect size).

^a = $p < .05$; ^b = $p < .001$

Table 4. Pairwise comparisons between simple and complex versions of tasks for all measures

Dependent variable	Narrative	Map	Fire chief
Errors per AS-unit	.248	.000 ^b	.586
Ratio errors/words	.001 ^a	.000 ^b	.586
Error-repairs per AS-unit	.128	.060	.198
Ratio error-repairs/words	.241	.000 ^b	.066
All repairs per AS-unit	.102	.020 ^a	.869
All repairs per 100 words	.115	.002 ^a	.001 ^b
Percentage of self-repairs	.283	.000 ^b	.065
Ratio of repaired to unrepaired errors	.035 ^a	.010 ^a	.070
Corrected ratio of repaired/unrepaired errors	.036 ^a	.005 ^a	.008 ^a

^a = $p < .05$; ^b = $p < .001$

in terms of the ratio of repaired to unrepaired errors or the proportion of self-repairs. The only unusual behavior was found when all repairs were computed together and divided by the number of words, the fire chief task displaying a lower rate than the map task under complex performance.

Table 5. Pairwise comparisons between task types under simple and complex performance

Task types	Simple			Complex		
	Narrative and Map Task	Narrative and Fire chief Task	Map Task and Fire chief Task	Narrative and Map Task	Narrative and Fire chief Task	Map Task and Fire chief Task
Errors per AS-unit	.359	.027 ^a	.188	.011 ^a	.743	.000 ^b
Ratio errors/words	.006 ^a	.036 ^a	.629	.114	.132	.013 ^a
Error-repairs per AS-unit	.000 ^b	.001 ^a	.021 ^a	.000 ^b	.001 ^a	.005 ^a
Ratio error-repairs/words	.000 ^b	.000 ^b	.138	.000 ^b	.001 ^a	.042 ^a
All repairs per AS-unit	.000 ^b	.602	.000 ^b	.000 ^a	.065	.009 ^a
Ratio all repairs to words	.000 ^b	.000 ^b	.000 ^b	.000 ^a	.001 ^a	.001 ^a
Percentage of self-repairs	.000 ^b	.006 ^a	.076	.001 ^a	.006 ^a	.182
Ratio of repaired to unrepaired errors	.001 ^a	.057	0.69	.004 ^a	.004 ^a	.158
Corrected ratio of repaired/unrepaired errors	.000 ^b	.010 ^a	.481	.037 ^a	.112	.661

^a = $p < .05$; ^b = $p < .001$

5.4. Results of Hypothesis 3: Performance between low and high proficiency groups

Except for error behavior, no significant differences were found between the high and low group with regard to self-repair behavior. Learners in the low proficiency group made more errors than learners in the high proficiency groups as may be expected. As for self-repair behavior, learners in the low proficiency group self-repaired more frequently in the complex version of the narrative task, as shown by the number of errors per AS-unit ($p < .05$) and the ratio of error to words ($p < .05$). Hypothesis 3 was therefore largely confirmed.

6. Discussion

6.1. Hypothesis 1

Hypothesis 1 stated that all tasks performed under complex conditions would trigger less errors and a significantly higher rate and proportion of repairs. This hypothesis therefore tried to check *within* tasks differences. As seen in the results section, this was only partially confirmed.

The narrative task performed in the past tense and without the visual presence of the comic strip was not perceived as more difficult than its simple counterpart but without reaching statistical significance. This contradicts Gilabert's (2005, 2007) findings which clearly showed the complex task to be more difficult than the simple version. In that study, however, the level of student was homogenous, while in the present study more advanced students participated. The fact that more advanced learners may not have perceived the performance of a task with a degree of displaced, past time reference as difficult may have affected the results of the group as a whole. With regard to the frequency of errors and self-repairs, a higher rate of errors was captured by one of the measures (ratio errors/words), but not by the number of error per AS-units. It could be argued that taking words as a reference makes the measure more sensitive, since differences in the length of AS units may distort the picture of how often learners made an error. Although they used different measures to calculate attention to form, this is in line with findings by Rahimpour (1997), Iwashita et al. (2001), and Ishikawa (2007) regarding accuracy in narrative performance. As for the amount of self-repair, while the percentage of self-repair did not capture any differences, significant differences were found when the ratio of repaired to unrepaired errors and its corrected version were calculated. This would suggest that increased Task Complexity may have led learners to notice more errors that needed to be repaired and repaired them. This matches recent findings by Gilabert (2005, 2007). The effect of complexity on self-repair behavior during narrative production in this study, however, only shows differences in proportion but not in the rate of self-repair. Two psycholinguistic explanations may be provided by taking Levelt's (1993) and Kormos' (2006) models of L1 and L2 production as a reference. Firstly, increases in Task Complexity caused by the visual absence of the comic strip may have triggered changes in macro-planning. As opposed to the Here-and-Now condition which lets students narrate the story picture by picture, the There-and-Now forces learners to keep the whole story in mind as they perform the task which, in turn, pushes them to achieve interpropositional coherence (Givon 1985). This, I would argue, may draw their attention to how certain events need to be connected and how they need to be expressed in the L2. In this context, monitoring plays the role of testing and revising the expression of such concepts. Secondly, Task Complexity may also cause changes in micro-planning. In the

There-and-then condition, learners were instructed to narrate the story in the past tense. Reference to the past tense can be made by means of lexical devices (i.e., *yesterday*, *last night*) or morphological ones, such as the use of *-ed* for regular verbs or irregular past tense forms. Although no qualitative analysis of the data was carried out, it could be speculated that learners may have been led by task design to focus on past tense morphology to maintain reference to the past as they were instructed. This may explain why increased monitoring took place. It is worth noting at this point that findings of the comparison of low and high proficiency groups showed that learners in the low proficiency group repaired more often in the complex task. It could be argued, then, that the operationalization of the variable may have captured the behavior of the low proficiency group more clearly since they can be expected to have more problems with past tense morphology, and suggests the need to test this variable with advanced learners alone.

As for the map task, the complex task was perceived as significantly more difficult than the simple version. This difference in perception was the strongest of all tasks. Both the rate and amount of self-repairs displayed significant differences when the task was complexified along the number of elements and axes learners needed to make reference to. On the one hand, learners made less errors, as captured by the measures of errors per AS-unit and ratio of errors to words. This confirms previous findings by Kuiken, Mos and Vedder (2005), Kuiken and Vedder (2007) and Michel (this volume). Both the rate and amount of self-repair were significantly affected by increases in task demands. Learners were pushed to specify referential location (e.g., “*and before the building no behind the building*”, “*well ok the fourth bank if you go to the from the by the stairs ok*”) by referring to a set of similar landmarks, and direction towards the goal (e.g., *going to with the stairs down you can see it*) and perspective (e.g., “*and then turn right and then at the back you’ll find the the pet shop to buy the food*”) along different axes. Additionally, since few street names were provided in the complex map and, as a consequence, their degree of similarity was higher, learners had to specify the precise references they were communicating (e.g., *so you have to go through the not to the avenue the big avenue*). Without a qualitative insight into the data, it may be argued that learners had to check such complex concepts against their receptive knowledge of the L2 in order to achieve precision (de Bot 1996: 551). When wrong or imprecise linguistic devices to express such abstract and complex concepts were selected, used, and detected, learners may have been forced to review their choices and resort to simpler or alternative descriptions available in their repertoire. In such a process, monitoring may have played a larger role than in the simple version of the task, where simple linguistic devices for simpler concepts may have been readily available. This interpretation points in the direction of the Cognition Hypothesis (Robinson 2001b, 2003, 2005, 2007, this volume) which predicts that

more complex concepts will draw learners' attention to the way they should be grammatized in the L2. As for the issue of the effects of limited attention on self-repairs, results for the map task cast a doubt on the argument that when task demands are increased, less attention may be devoted to self-repairs. In my view, the claim may hold true when tasks are made more demanding along resource-dispersing variable but not along resource-directing ones.

Finally, the decision-making fire chief task also triggered differences in the perception of difficulty, the complex task being perceived as more difficult than its simple counterpart. When engaged in complex performance, the fire chief task was perceived as the most difficulty (i.e., significantly more than the narrative and with no significant differences as compared to the map task) and learners felt the least confident in their performance (i.e., significantly less than in the other two tasks). As for errors and repairs, it generated similar a number of errors and repairs during both simple and complex performance, with one frequency measure and one measure of the amount of self-repair pointing in the direction of the hypothesis. With limited pre-task planning time, in the complex version of the task the cognitive/conceptual demands of the task were increased by leading learners to establish complex connections between events and prioritize them. Complex cause-effect relationships which also required learners to determine the mental states of the characters in the task forced multiple-step decisions as opposed to single or limited step decisions in the simple version of the task. This may have triggered revision not only of macro- and micro plans but also of the overt speech used to express such complex decisions, as captured by the corrected ratio of repaired to unrepaired errors. The effect of Task Complexity on self-repairs in the complex decision-making task is, however, not as strong as in the two other tasks. As will be argued in the next section, given that no pre-task planning time was provided, learners may have geared their attention towards the justification of their actions, with more resources being employed on structural complexity. If we consider perception results and the fact that no planning time was provided, this may suggest that trying to decide on the actions to be taken, sequencing them in a reasonable order, and justifying such sequence may have focused learners on conceptualizing messages and, as a consequence, it may have taken a toll on their monitoring.

It can therefore be concluded that increasing tasks demands along resource-directing dimensions has the potential to draw attention to how messages are being encoded during performance. This was more obvious for the map tasks than the narrative or the decision-making task, which showed weaker levels of significance. A more detailed study of the quality of self-repairs in this type of task should look into whether there exists a specific pattern of self-repair behavior.

6.2. *Hypothesis 2*

Hypothesis 2 stated that, when compared, the narrative and map task would trigger enhanced self-repair behavior as compared to the fire chief task. This hypothesis therefore tried to capture *between* tasks differences. As seen in the results section, this was only partially confirmed. It was argued that because the narrative provides a very clear storyline and the map task a marked route, fewer resources should be used for the generation of communicative intention. In terms of perception, the narrative task was construed as significantly simpler than the decision-making task under simple conditions, and simpler than both the map task and the fire chief task under complex performance. The map task and the fire chief task, however, did not display any significant differences in perceptions of task difficulty. Despite having a clear route that learners had to instruct their friend to follow, the operationalization with many similar elements and along different axes may have led them to perceive the map task as significantly more difficult than the narrative. As for errors and self-repair behavior, the narrative task triggered a higher rate of errors than the fire chief task in the simple version, and more than the map task under complex performance. However, in general, the narrative triggered the highest rate and amount of self-repairs of the three tasks. Two explanations may be provided: one, more resources were available for monitoring given the lesser demands along resource dispersing variables (i.e., not having to create the storyline from scratch and not having to establish multiple-step connections between events); two, text length in the narrative was much shorter and measures may have been distorted. The corrected ratio of repaired to unrepaired error, however, may disprove this argument. It can then be concluded that for the narrative task the prediction was confirmed. When we compare the map task to the fire chief task the picture is quite different. Learners made proportionally less errors in the map task but did not self-repair more than in the fire chief task under either simple or complex performance. This goes against the prediction, which suggested that because in the fire chief task communicative intention has to be generated from scratch, resources would be deviated from monitoring. Without significant differences, the map task triggered less self-repairs than the fire chief task. I would argue that in the fire chief attention may have been allocated to the argumentation to justify the choice of actions and their sequence, which may have drawn their attention to more complex syntactic structures. If competition for attention existed, it may have fallen on the side of complexity rather than accuracy. The question remains of how learners allocated their attention during the performance of the map task, which showed a higher percentage of repairs in the complex versions but at the same time triggered less self-repairs than the other two tasks (only significantly when compared to the narrative task). In the map task, few street names were provided so that learners would have to make ref-

erences to other landmarks. This may have drawn their attention to the words used to express the different landmarks (e.g., *building*, or *parking lot*) and less to self repairs, suggesting competition between accuracy and lexical complexity. These predictions, however, need to be tested and should motivate further research.

6.3. Hypothesis 3

Results of comparing low and high proficiency groups are in line with previous findings in the self-repair literature (see Kormos 2006: 133). Students in the low proficiency group displayed a higher number of errors but did not repair more often than their high-proficiency counterparts. As mentioned before, only the measures of error-repairs per AS-units and the ratio of error-repairs to words suggest that learners needed to repair more frequently when narrated in the past. This may be explained by the fact that students in the high proficiency groups may not have encountered problems with verb morphology. The results in this study confirm Kormos' (2006: 133) claim that: "... with increasing L2 proficiency there is a shift from simple error repairs to more complex discourse-level repairs, but the global frequency of self-corrections does not seem to be affected by the level of L2 competence."

7. Conclusions

As an answer to the first question advanced in this study, I would propose that when task demands were increased along resource-directing variables learners' attention was drawn to the way they were encoding their messages, with positive effects for accuracy. I agree with Kormos (2006) when she suggests that higher tasks demands may deviate attention from monitoring. For example, because in this study no pre-task planning time was provided, maintaining fluency and building complex structures may have prevented learners from monitoring more. However, results of the three tasks either show a strong trend or a significant effect of Task Complexity on self-repair behavior when they were made more complex along resource-directing variables. This happened at different levels in different task types in which different variables were manipulated. Attention to form, as shown by self-generated repairs, happens during performance and task design can further encourage it. As argued in Section 3.1, this has the potential to foster learners's acquisition of the L2 by having them revise their hypotheses about the L2, notice gaps in their interlanguage, and push their interlanguage development. As for the second question, there are differences in how learners tackle different task types as shown by different self-repair behaviors. This confirms what has already been found in previous self-repair and Task Complexity studies and suggests that future research should keep exploring and comparing the impact of manipulating Task Complexity on different

task types. Certainly, subsequent studies on the same data should look at how accuracy interacts with the dimensions of fluency and lexical and structural complexity. It remains an issue how task design may force learners to make strategic decisions, in Wickens' (this volume) terms, as to which dimension of production (i.e., fluency, complexity, or accuracy) they allocate attention to. This is particularly interesting if tasks can 'train' learners to focus on certain dimensions either individually or simultaneously. I would also suggest that the operationalization of variables would benefit from further discussion and fine-tuning. Finally, although proficiency did not affect self-repair behavior, future research into other individual differences such as aptitude, intelligence, and working memory capacity should shed light on how learners monitor their L2 speech.

Finally, this study has constituted an attempt at measuring the performance of the same learners with three different task types. By focusing on self-repair behavior, it has shown how tasks can be manipulated along the cognitive demands they impose on learners and how these may draw learners' attention to form. This has important implications for task design as devised for classroom performance and testing. Results have also suggested that the same learners behave differently when performing different tasks types and that obtaining information about how learners perform is paramount for design and sequencing decisions in language learning programs.

Universitat Ramon Llull
<RogerGG@blanquerna.url.edu>

Notes

- * Special thanks to the members of the Language Acquisition Research Group at the University of Barcelona and to J. Barón and M. A. Llanes for making this research possible.
- 1. See also Dörnyei and Kormos (1999) for a detailed classification of problem-solving mechanisms.
- 2. Phonological repairs have been left out of the count since there was no agreement in determining what constitutes a phonological error.
- 3. Piloting of the experimental was carried out with 25 students before data collection in order to test their design and make sure that students' perception matched the different operationalizations of Task Complexity in different tasks (Montreal, 2006). 2 comic strips used and tested in Gilabert (2005) were selected. The maps used in the first pilot did not seem to distinguish the two levels of complexity and were re-designed and piloted again. The opinion-giving task was also changed after the pilot study and piloted a second time until a distinction in perception of difficulty was achieved.
- 4. The definition of error is adopted from Lennon (1991: 182): "a linguistic form or combination of forms, which in the same context and under similar conditions of

- production would, in all likelihood, not be produced by the speakers' native speaker counterparts."
5. Phonological self-repairs were not considered either, even if they are relatively easy to detect. The reason is to be found in the difficulty to reach an agreement between raters as to what constitutes a phonological error.
 6. Although .05 is the standard α in the literature, it was decided to follow the suggestion by one of the reviewers of lowering the level of significance by applying the Bonferroni correction. Such a correction is applied when multiple hypotheses are tested on the data in order to avoid finding significance levels by chance. The Bonferroni correction therefore triggers much more conservative levels of significance and it is a highly controversial issue in the literature (See Perneger 1998 for a discussion).
 7. Interrater reliability was calculated by means of percentage agreement.

References

- Bédart, Serge (1991). Word interruption in self-repairing. *Journal of Psycholinguistic Research* 20: 123–137.
- Chown, Eric, Stephen Kaplan and David Kortenkamp (1995). Prototypes, location and associative networks (PLAN): Towards a unified theory of cognitive mapping. *Cognitive Science* 19: 1–52.
- Cornell, Edward, Donald Heth and Denise Alberts (1994). Place recognition and way finding by children and adults. *Memory and Cognition* 22: 633–643.
- de Bot, Kees (1996). The psycholinguistics of the output hypothesis. *Language Learning* 46: 529–55.
- Dörnyei, Zoltan and Judit Kormos (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition* 20: 349–385.
- Ellis, Rod (ed.) (2005). *Planning and Task Performance in a Second Language*. Amsterdam: John Benjamins.
- Evans, M. W. and J. Marciniak (1987). *Software Quality Assurance and Management*. New York: John Wiley & Sons.
- Foster, Pauline and Peter Skehan (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition* 18: 299–323.
- Foster, Pauline, Alan Tonkyn and Gillian Wigglesworth (2000). Measuring spoken discourse: A unit for all reasons. *Applied Linguistics* 21: 333–353.
- Gass, Susan and Alison Mackey (2007). Input, interaction and output: An overview. *AILA Review* 19: 3–17.
- Gilabert, Roger (2005). *Task complexity and L2 narrative oral production*. Unpublished Ph. D. dissertation. University of Barcelona, Spain.
- (2006). Manipulating task complexity across task types: Effects on oral production. Paper presented at the AAAL/CAAL language production symposium in Montreal, Canada.
- (2007). The simultaneous manipulation of task complexity along planning and +/- Here-and-Now: effects on L2 oral production. In *Investigating Tasks in Formal Language Learning*, María del Pilar García-Mayo (ed.), 44–68. Clevedon: Multilingual Matters.
- Givón, Talmy (1985). Function, structure, and language acquisition. In *The Crosslinguistic Study of Language Acquisition*, Vol. 1, Dan Slobin (ed.), 1008–1025. Hillsdale, NJ: Lawrence Erlbaum.
- Harrington, Michael and Mark Sawyer (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition* 14: 25–38.

- Ishikawa, Tomohito (2007). The effects of manipulating task complexity along the +/-Here-and now in L2 written narrative discourse. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García-Mayo (ed.), 136–156. Clevedon: Multilingual Matters.
- Iwashita, Noriko, Tim McNamara and Catherine Elder (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.
- Kormos, Judit (1998). A new psycholinguistic taxonomy of self-repairs in L2: A qualitative analysis with retrospection. *Even Yearbook, ELITE SEAS Working Papers in Linguistics* 3: 43–68.
- (1999). Monitoring and self-repair in L2. *Language Learning* 49: 303–342.
- (2000). The role of attention in monitoring second language speech production. *Language Learning* 50: 343–384
- (2006) *Speech Production and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Kuiken, Folkert, Maria Mos and Ineke Vedder (2005). Cognitive task complexity and second language writing performance. In *EUROSLA Yearbook 5*, Susan Foster-Cohen, Maria del Pilar García-Mayo and Jasone Cenoz (eds.), 195–222. Amsterdam: John Benjamins.
- Kuiken, Folkert and Ineke Vedder (2007). Cognitive task complexity and linguistic performance in French L2 writing. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García-Mayo (ed.), 117–135. Clevedon: Multilingual Matters.
- Lennon, Paul (1991). Error: Some problems of definition, identification and distinction. *Applied Linguistics* 12: 180–195.
- Levelt, Willem J. M. (1983). Monitoring and self-repair in speech. *Cognition* 33: 41–103.
- (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- (1993). Language use in normal speakers and its disorders. In *Linguistic Disorders and Pathologies*, Gerhard Blanken, Jurgen Dittmann, Hannelore Grimm, Jim Marshall and Claus Wallesch (eds.), 1–15. Berlin: Mouton De Gruyter.
- Long, Michael (2000). Focus on form in task-based language teaching. In *Language Policy and Pedagogy*, Richard Lambart and Elana Shohamy (eds.), 179–192. Amsterdam/Philadelphia: John Benjamins.
- Lyster, Roy and Leila Ranta (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition* 19: 37–66.
- MacWhinney, Brian (1995). *The CHILDES project: Tools for Analyzing Talk* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.
- Meara, Paul and Jim Milton (2003). *X Lex: The Swansea Vocabulary Levels Test*. Express, Newbury.
- Niwa, Yukiko (2000). Reasoning demands of L2 tasks and L2 narrative production: Effects of individual differences in working memory, intelligence, and aptitude. Unpublished M.A. dissertation, Aoyama Gakuin University, Tokyo.
- Nuevo, Ana-Maria (2006). Task complexity and interaction. Unpublished Ph. D. dissertation, Georgetown University, U.S.A.
- Perneger, Thomas (1998). What is wrong with Bonferroni adjustments. *British Medical Journal* 316: 1236–1238.
- Quesada, Jose, Walter Kintsch and Emilio Gómez-Millán (2005). Complex problem-solving: A field in search of a definition? *Theoretical Issues in Ergonomic Science* 6 (1): 5–33.
- Theory of complex problem solving using the vector space model (part 2). Latent Semantic Analysis applied to empirical results from adaptation experiments. *Cognitive Research with Microworlds* 2: 147–158.
- Rahimpour, Massoud (1997). Task condition, task complexity and variation in L2 discourse. Unpublished Ph. D. dissertation, University of Queensland, Australia.
- Lee, Eliza and Leslie Rescorla (2002). The use of psychological state terms by late talkers at age 3. *Applied Psycholinguistics* 23: 623–641.

- Robinson, Peter (1995a). Task complexity and second language narrative discourse. *Language Learning* 45: 99–140.
- (1995b). Attention, memory and the 'noticing' hypothesis. *Language Learning* 45: 283–331.
- (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22: 27–57.
- (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In *Cognition and Second Language Instruction*, Peter Robinson (ed.), 287–318. Cambridge: Cambridge University Press.
- (ed.) (2002). *Individual Differences and Instructed Language Learning*. Philadelphia: John Benjamins.
- (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies* 21: 45–105
- (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics* 45: 1–32.
- (2007). Criteria for classifying and sequencing pedagogic tasks. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García-Mayo (ed.), 7–27. Clevedon: Multilingual Matters.
- Schegloff, Emanuel, Gail Jefferson and Harvey Sachs (1977). The preference of self-correction in the organization of repair in conversation. *Language* 53: 361–382.
- Shatz, Marilyn, Wellman, Henry and Susan Silber (1983). The acquisition of mental verbs: A systematic investigation of first references to mental states. *Cognition* 14: 301–321.
- Skehan, Peter (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan, Peter and Pauline Foster (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research* 1: 185–211.
- (2001). Cognition and tasks. In *Cognition and Second Language Instruction*, Peter Robinson (ed.), 183–205. Cambridge: Cambridge University Press.
- Slobin, Dan (1985). Crosslinguistic evidence for the language-making capacity. In *The Crosslinguistic Study of Language Acquisition*, Vol. 2, *Theoretical Issues*, Dan Slobin (ed.), 1157–1259. Hillsdale, NJ: Erlbaum.
- Swain, Merrill (1998). Focus on form through conscious reflection. In *Focus on Form in Classroom Second Language Acquisition*, Catherine Doughty and Jessica Williams (eds.), 64–81. New York: Cambridge University Press
- Talmy, Len (2000). *Toward a Cognitive Semantics*, Vol. 1, *Concept Structuring Systems*. Cambridge, MA: MIT Press.
- Tomlin, Russ and Victor Villa (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition* 16: 185–204.
- Van Hest, Erna (1986). *Self-repair in L1 and L2 Production*. Tilburg, Netherlands: Tilburg University Press.
- Wickens, Christopher D. (1989). Attention and skilled performance. In *Humans Skills*, 2nd edition, Daniel Holding (ed.), 71–105. New York: John Wiley.
- (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins.
- Yuan, Fang and Rod Ellis (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics* 24: 1–27.

Copyright of IRAL: International Review of Applied Linguistics in Language Teaching is the property of Walter de Gruyter GmbH & Co. KG. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.