

Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance¹

ROGER GILABERT, JÚLIA BARÓN AND ÀNGELS LLANES

Abstract

The goal of this study is to investigate the impact of manipulating the cognitive complexity of three different types of oral tasks on interaction. The study first considers the concepts of task complexity and interaction and then examines the specific studies that have looked at the effects of increasing task complexity on conversational interaction. In the experiment, learners of English as a foreign language organized into 27 dyads carry out three different types of tasks: a narrative reconstruction task, an instruction-giving map task, and a decision-making task. Two different versions of each task, one simple and one complex, are presented to learners in different sequences. Task complexity is manipulated along the degree of displaced, past time reference, the number of elements, and the reasoning demands. Audio recordings are transcribed and coded for interactional feedback, which is measured in terms of negotiation of meaning (i.e., confirmation checks, clarification requests, and comprehension checks), recasts, language-related episodes (LREs), and repairs, all of which have been described in the literature as being conducive to acquisition. Both parametric and non-parametric statistical tests are used. Results are discussed in the light of previous studies that have looked at the specific relationship between task complexity and interaction, attention models (Robinson 2001a, 2003, 2005, 2007b; Skehan and Foster 2001), and how different task types may variously affect the way interaction proceeds during task performance.

1. Acknowledgments: Special thanks to the members of the Language Acquisition Research Group (GRAL) at the University of Barcelona. This research would not have been possible without the help of the MEDU (Ministerio de Ciencia y Tecnología – 519773) and the AGAUR research agency from Generalitat de Catalunya (L 2006 ARIE10071).

1. Introduction

Most researchers, syllabus designers, and teachers who embrace tasks as units of pedagogic intervention would agree that giving learners tasks so that they can put their language to use and interact serves the purpose of preparing them for real-life communication outside the classroom. In the last three decades, there has been a growing interest in how the manipulation of internal features of tasks may affect L2 learners' production and acquisition, which are seen as the ultimate goals of communicative language approaches such as task-based learning teaching, project work, and language and content integrated learning. From a cognitive perspective, researchers have tried to tap into the effects that manipulating certain task features may have on learners' fluency, structural and lexical complexity, and accuracy. As for the interaction agenda, one of its main objectives has been to measure how task design affects the interactional moves that take place between or among learners, between learners and native speakers, and between learners and teachers, and how these may contribute to language development. This has been done by manipulating task characteristics such as the flow of information, information distribution, the convergence or divergence of task goals, or the limited or unlimited solutions to the task. Conversational moves triggered by such manipulations may involve requesting clarification, confirming information, confirming comprehension, providing corrective feedback to the interlocutors' wrong productions, or negotiating the language involved in the message, and are said to help learners in the acquisition of the target language. As we will see in the next section, the Cognition Hypothesis (Robinson 2001a, 2001b, 2003, 2005, 2007a, 2007b; Robinson and Gilabert 2007) and its associated Triadic Componential Framework for pedagogic task classification and task design bring together the cognitive and interactive research agendas and provide a theoretical ground to analyze and explain how L2 production and acquisition may be fostered by task design.

Although a number of studies have started to explore the synergies existing between different degrees of internal complexity of tasks and its effects on interaction (Robinson 2001a, 2007b; Nuevo 2006; Michel et al. 2007; Révész 2007), and other studies have checked the impact of interaction across task types (Gass et al. 2005), no study to date has looked at whether the impact of increasing task complexity on interaction is the same across different varieties of tasks. Within the framework of the Cognition Hypothesis, the goal of this paper is to measure the effects of increasing cognitive task complexity on learner-learner interaction across task types. In order to do so, an experiment was designed so that the same learners would carry out a simple and a complex version of three different task types (i.e. a narrative reconstruction task, an instruction-giving task, and a decision-making task), hence performing a total

of six tasks in pairs, which were used to elicit oral data that was subsequently used to measure the amount of interaction.

1.1. *Task complexity and the Cognition Hypothesis*

The Cognition Hypothesis of adult task-based language learning (Robinson 2001a, 2001b, 2003, 2005, 2007a, 2007b) establishes the existence of a very strong link between the cognitive load tasks impose on learners' processing and their production and development. This is based on findings from both functional/cognitive linguistics and L1 developmental psychology. As tasks are made more complex, learners need to stretch their attentional, memory, reasoning, and other information processing resources in order to meet the cognitive demands imposed by task design on their processing. This, Robinson suggests, has the potential to draw learners' attention to a wider range of vocabulary (lexical complexity), to focus on the way they grammaticize concepts (accuracy) and how they syntacticize them (structural complexity), and to increase interaction. Specifically with regard to interaction, Robinson (2003) claims that increasing task cognitive demands will generate more communication breakdowns. These, in turn, will provide more opportunities for learning by generating conversational episodes that will create the conditions for noticing and uptake of more salient input. As we will see in the next section, saliency, noticing, and uptake have been claimed to be facilitated by conversational episodes such as clarification requests, confirmation checks, comprehension checks, recasts, and language-related episodes. While for monologic production Robinson suggests that his predictions apply when tasks are manipulated along resource-directing dimensions (e.g., the number of elements in a task, the amount of reasoning required from learners, the degree of displaced time reference), for dialogic tasks his predictions also subscribe to tasks that have been complexified along resource-dispersing dimensions (e.g., by giving less planning time, making them less familiar, or making them dual tasks). The distinction between resource-dispersing factors, which deviate attentional and memory resources from processing during task performance, and resource-directing ones, which direct learners' attention to the way their messages are being encoded, is paramount for the predictions made by Robinson (2001a, 2001b, 2005, 2007a, 2007b). This is especially so when there are divergent positions as to the way attention is allocated during task performance. Some researchers conceive attention as being a capacity-limited single pool of resources (Skehan 1998; Skehan and Foster 1997, 2001). For Robinson, who bases his stance on the work of researchers such as Allport (1987), Gopher (1992), and Wickens (1989, 1992, 2007), attention may draw on various pools during performance, which may explain why under certain conditions

(e.g., with tasks being simple along resource-dispersing resources and complex along resource-directing ones) attention can be devoted to both the meaning (lexical complexity) and form (accuracy) of utterances. This paper focuses on the claims about the impact of task complexity on learner interaction by manipulating resource-directing variables.

1.2. *Interaction*

In general, numerous researchers have defended the idea that engaging in interaction is beneficial for interlanguage development. The main theoretical arguments used to defend the role of interaction in acquisition can be drawn from both the Interaction Hypothesis (Gass 1997, 2003; Gass and Mackey 2007; Long 1996; Pica 1994) and the Output Hypothesis (Swain 1995, 2005). Researchers such as Long (1996), Gass (1997) and Pica (1994), for example, have suggested that engaging in interaction may lead to communication breakdowns that lead to negotiation of meaning. In a negotiation episode, learners typically get more comprehensible input and negative feedback. More specifically, negotiation causes rearrangement, segmentation, and movement of sentence constituents that provide better comprehensibility and make input and output more salient. Hence negotiation in which clarifications are requested or the accuracy of the information is confirmed may lead to more appropriate or more accurate output (Lyster 1998; McDonough 2005; Pica 1994). Another function has been identified as the “noticing/triggering” function of engaging in interaction. Following Schmidt’s (1990) concept of “noticing the gap”, Swain suggests that students may notice the gaps between what they want to say and what they can actually produce, therefore realizing the knowledge they lack about the target language. This recognition of gaps in knowledge can be internal, so that it is the student himself or herself that notices the gap, or external, that is, detected by the interlocutor or teacher. If input is immediately available, as may be the case in interaction, students might pay more attention to subsequent input in order to solve their problem and fill their gap. Additionally, Gass (1997) subscribes to Swain’s (1995, 2005) suggestion of output as a generator of hypotheses, and the idea of output as a generator of feedback that can be used to accept or reject such hypotheses. By receiving feedback from their interlocutors, learners engage in negotiation of meaning, and their production is pushed to produce more accurate and precise language. Another argument is the metalinguistic function of engaging in interaction, understood as using linguistic terminology to think about language. According to Swain (1995, 2005), through metatalk, learners engage in syntactic processing that goes beyond the need to be understood. In a genuine communicative context, thinking about rules, forms, and form-function relationships shows “language in progress”, and may lead learn-

ers to produce modified or reprocessed output which in turn may contribute to language acquisition. Other arguments such as the automatization (Gass 1997; Lyster and Ranta 1997; Skehan 1998) function of interaction and discourse skills development (Skehan 1998) have also been advanced as benefits of interactive performance.

As for the empirical studies taking the Interaction Hypothesis to test, the first studies (see Mackey and Jaemyung Goo [2007] for a thorough review) focused on the relationship of negotiation and comprehensible input, which lead research to focus on conversational episodes such as clarification requests, confirmation checks, and comprehensions checks, all said to be conducive to acquisition. Progressively, the interest turned to language-related episodes and recasts. Many researchers (Swain and Lapkin 1995, 1998, 2001; Swain 1995, 1998; Lyster and Ranta 1997; Lyster 1998; Alegría de la Colina and García-Mayo 2007) have claimed that when learners engage in language-related episodes development is fostered because they check and question their hypotheses about the target language, they notice holes in their interlanguage, they notice certain forms in their interlocutor's speech by repeating them, they self-repair, and they engage in metalinguistic talk. The literature on recasts has probably provided the most robust evidence connecting interaction and acquisition. Some examples are the direct association of interaction and past tense acquisition (Han 2002; McDonough 2007), interaction and the acquisition of verbal morphology (Ishida 2004; Iwashita 2003), and question development (Mackey and Philp 1998). This paper assumes that interaction is associated with language development (Gass and Mackey 2007). As for recasts, it assumes that these are beneficial for interlanguage development, but it specifically avoids the debate on the effectiveness of recasts, their immediate or delayed effect on development, or the impact of different learners' responses to recasts (e.g., elicitations vs. recasts) on development. For that debate see McDonough and Mackey (2006) and Nassaji (2007). This study therefore assumes the basic claims of the benefits of interaction based on the extensive, although still somewhat inconclusive, literature on interactional feedback.

1.3. *Task complexity and interaction*

As opposed to monologic production, dialogic production is still under-researched within the framework of the Cognition Hypothesis. To our knowledge, only a handful of studies have looked at the specific relationship between task complexity and interaction. For the sake of space, this brief review only reports on the hypotheses and findings related to the specific effects of task complexity on interaction, even if the studies looked at other features. Robinson (2001a) operationalized the concept of task complexity in one of the few

studies that combined two variables, familiarity and the number of elements, simultaneously. Robinson used an interactive one-way map task with which one student was to give directions to another student who had to draw a route on an empty map. The simple version included few elements and references of a small area which was also known to the students, while the complex map consisted of a large area with many elements and that was unknown to the students. Specifically in terms of interaction, he hypothesized that the most complex version of the task would trigger more clarification requests, confirmation checks, and comprehension checks. He found a significantly higher number of comprehension checks and a strong trend for more clarification requests in the complex version. Nuevo (2006) manipulated complexity with regard to the structure of a narrative task. In the simple version, the vignettes of the strip were ordered and only required the learners to narrate the story. In the complex version, learners were given unordered vignettes that they needed to rearrange as they narrated the story. Nuevo found task complexity to have an impact on confirmation checks and comprehension checks, and also on self-repairs and hypothesis testing, two sub-categories within language-related episodes. In a photo description task, Révész (2007) looked at the impact of recasts on language development as mediated by task complexity. She manipulated task complexity by either providing or removing visual support. In the simple version of the tasks, learners could look at the pictures while performing the tasks, while the pictures were removed from their sight when performing under complex conditions. As for the mediation of complexity in learning subsequent to recasting, Révész found recasts to be more effective for learning when produced under complex conditions. Michel et al. (2007) manipulated the number of elements in a product-description task. In the simple version, learners had to consider a few features while in the complex version they had to consider many features in order to complete the task. Although Michel et al. did not specifically measure interaction, she showed task complexity not to have an impact on self-repairs (a sub-category of LREs in the present study) in her dialogic tasks. Finally, Robinson (2007b) operationalized task complexity by increasing the reasoning demands of a picture description task. By using pictures from the Wechsler Adult Intelligence Scale-Revised (WAIS-R) test, Robinson established three levels of increasing complexity, from low intentional reasoning demands to high intentional reasoning demands. Robinson's results showed that increasing task complexity progressively promoted the use of clarification requests and confirmation checks, hence confirming one of the predictions of the Cognition Hypothesis.

1.4. *Questions and hypotheses*

Based on findings from both the Task Complexity and the Interaction literature, the general questions that motivate this study are:

1. Does increasing the cognitive complexity of tasks generate more learner-learner interaction?
2. Are the effects of increasing task complexity the same across task types?

Following the predictions of the Cognition Hypothesis (Robinson 2001a, 2001b, 2003, 2005, 2007a, 2007b) we can formulate the following hypotheses:

Hypothesis 1 Increasing the cognitive complexity across tasks will generate more negotiation of meaning (i.e., clarification requests, confirmation checks, and comprehension checks), more LREs and more recasts.

Hypothesis 2 There will be no differences in the impact of task complexity across the different task types.

The first hypothesis is motivated both by the prediction of the Cognition Hypothesis and by the findings from a few studies (Robinson 2001a, 2007b; Nuevo 2006) which have indicated that Task Complexity has an impact on interactional moves. As for the second hypothesis, the null hypothesis applies since we have no previous conclusive evidence as to the differential impact of different task types on interaction as mediated by task complexity that may motivate any hypothesis in any direction.

2. Experimental design

A repeated-measures design was used in which learners performed the six tasks in six different sequences (e.g., simple narrative > complex narrative > simple instruction-giving task > complex instruction-giving task > simple decision-making task > complex decision-making task), the within-learner factor being Task Complexity. Besides task type, the sequence of conditions under which the tasks were performed was also thought to potentially affect performance. Students were therefore randomly assigned to one of the six groups in a Latin Squares design. This was done to prevent any practice or carryover effects that may occur from one task to another. To measure learners' perception of task complexity, repeated measures analyses of variance (ANOVA) of the affective variables (difficulty, stress, confidence, interest, and motivation) were carried out. 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 created by Robinson (2001a).

2.1. *Participants*

Sixty volunteers from the English Studies department at the University of Barcelona participated in the study. They were all taking an English major at university. Students were given extra credit for their participation. Students had received instruction in English for approximately the same number of years. Learners' ages ranged between 18 and 40. X-lex and Y-lex placement tests (devised, developed, and widely tested by Meara and Milton [2003] at the University of Wales in Swansea), which measure vocabulary size, were used to control for potential differences in proficiency. Out of the original 30 dyads, only 27 dyads were finally included in the study, since two of the dyads were missing information and one of the dyads had a level of proficiency clearly above the average.

2.2. *Materials*

The tasks described below were used in a monologic fashion in Gilabert (2007c) and adapted to dialogic production in this study. In terms of their informational and interactive features (Pica et al. 1993) the three tasks used in this research were convergent, split information-gap tasks. Although all of them were interactive, the instruction-giving map task was one-way, with one speaker holding the information about the route to follow, while in the other two there was a two-way information flow (information was equally distributed between the two participants). The narrative and the instruction-giving task were closed (i.e., only one solution or a limited set of solutions are possible), while the opinion-giving task is open by nature (i.e. any solution is possible).

Narrative: Previously tested in Gilabert (2005) and further described in Gilabert (2007c), two wordless comic strips were employed to elicit narrative discourse. Learners were given a similar number of vignettes each and were asked to reconstruct the story. Regarding the +/– Here-and-Now distinction, this research adopted Robinson's (1995) operationalization. They were prompted to narrate in the present and while looking at the strips in the Here-and-Now version. For the There-and-Then version, participants were prompted to narrate the story in the past and the comic strips were removed from their sight during task performance.

*Instruction-giving task:*² In the simple version, one learner was prompted to ask a number of favors from his/her friend (the listener) which involved giving directions based on a route-marked city map to the other one, who did not have the route marked on the same map. The instructions involved describing how to get to a news stand for the listener to buy a newspaper, to the post office to post a letter, and to a flower shop to purchase a bunch of flowers before going back to a subway station. To do so, they had to ask their listener to move along a single lateral axis (i.e., left, right, straight). The instructions for the complex version included asking the listener to pick up their dog from the vet's and to buy food from a department store before going back to the subway station. This time, however, the task was manipulated to include many and similar points of reference. They had to ask their listener to move along the lateral (i.e. left, right, straight along the street), vertical (i.e., up and down stairs inside buildings), and sagittal axes (i.e., from the front to the back of the store). This operationalization was based on Chown et al. (1995), Cornell et al. (1994) and their constructs regarding landmark identification, path selection, direction selection, and abstract environmental overviews.

Decision-making task: In this task, learners were given a drawing of a building where a fire has broken out and a number of people need to be helped by rescue teams. They were asked to act as fire chiefs and were given slightly different information. Learners were prompted to describe the actions they would take, decide on the sequence of those actions, and justify their decisions as to actions and sequence. The task is an adaptation from the 'Fire chief' task used in cognitive psychology in which, as Quesada et al. (2001) suggest, complexity is associated with the intricacy of systems. Complexity was increased so that the connection between the variables involved taking a series of decisions in which early decisions may condition later ones. In the simple version, people in the building had no particular roles, learners had many resources (i.e., three fire trucks and a helicopter which could be used simultaneously) and few unrelated factors to consider (i.e., people in safe places, the fire being static, the smoke blowing away from the building). In the complex version, learners had to deal with specific types of people (e.g., a an elderly man, a pregnant woman with children, a severely injured person), which was thought would push them to

2. Piloting of the experiment was carried out with 20 pairs 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 (Gilabert 2007b). Two comic strips used and tested in Gilabert (2005) were selected. The opinion-giving task was modified so that learners would have different information, since in the original design both learners shared exactly the same information and this was thought to affect interaction by making the opinion-giving task different from the narrative and the map task.

refer to psychological state terms which in L1 acquisition are known to emerge in the order physiological > emotional > desire > cognitive state terms (Lee and Rescorla 2002; Shatz et al. 1983). In order to force the learners to prioritize and justify their actions, learners were given fewer resources (i.e. a single fire truck) and the factors in the task were closely related and dynamic (e.g., fires moving towards the people and smoke blowing into the building through the ventilation system) and all of them were designed to increase the cognitive demands of the task.

2.3. Procedures

Data collection took place in a single one-and-a-half hour session. Learners sat facing each other. First, the researcher created rapport with learners, then gave the instructions for the first task, left the room, and only came back to give instructions at the beginning of each of the other five tasks. In order to isolate task complexity from the influence of planning time, they were given a very short pre-task planning time (up to one minute before starting to perform). The steps taken during the session include a brief explanation of data collection procedures, small talk to establish rapport, and personal information retrieval. An affective variables questionnaire measuring perception of difficulty, stress, confidence, interest, and motivation was administered after each set of two (simple and complex) tasks.

2.3.1. *Measures.* In the study measures suggested by Gass et al. (2005) were adopted. These measures have been widely used in interactionist research and distinguish between (a) negotiation of meaning (i.e. clarification requests, confirmation checks, comprehension checks); (b) recasts; (c) and language-related episodes (LREs), on the basis of which information about self-repairs was calculated.

Following Long (1983: 137), clarification requests are any expression "... designed to elicit clarification of the interlocutor's preceding utterance(s)". For example:

Learner 1: go walking it's two apples further two streets more it looks

Learner 2: *two what?* (clarification request)

Learner 1: two streets further

In this conversational episode, learner 1 has incorrectly used the word "apples" to refer to "blocks" (in peninsular Spanish street blocks are "manzanas"), which are then replaced by "streets". Learner 2 needs clarification of Learner 1's utterance in order to understand what she needs to do.

As Long (1983:137) defines them, confirmation checks are any expressions “immediately following an utterance by the interlocutor which are designed to elicit confirmation that the utterance has been correctly heard or understood by the speaker”. For example:

Learner 1: ok it's in the it's in the corner the building

Learner 2: *in the corner?* (confirmation check)

Learner 1: yeah

In this interactional episode the learner tries to confirm whether the information she heard was right.

Comprehension checks, as Long (1983: 136) suggests, try to “anticipate and prevent a breakdown in communication”. For example:

Learner 1: I need that you go to get something in the first one that is the one for dogs, *yeah? you understand?* (comprehension check)

Learner 2: yeah dogs

In this interactional episode, the learner has been talking about three different pet shops and makes sure she has been understood appropriately.

A recast, as Nichols et al. (2001: 721) define it, is “a correct restatement of a learner's incorrectly formed utterance”. For example:

Learner 1: when you ah no no no ah ok so you ok you take the elevator and then you go out and then there are three ## and is in the right hand

Learner 2: ah ok is *in the right hand side* (REC) ok and I pick up the dog there

As for language-related episodes, we also adopt the definition provided by Swain and Lapkin (1998: 70), which suggests that they are “any part of a dialogue in which students talk about the language that they are producing, question their language use, or other- or self-correct”. For example:

Learner 1: when uh the first uh picture is when uh there are one *two pair un matrimonio? uh*

Learner 2: *a marriage?*

Learner 1: *yeah*

Learner 2: *a couple*

Learner 1: *a couple yes* (LRE)

In this conversational sequence, Learner 1 notices a gap in her interlanguage (i.e. she does not know the meaning of ‘married couple’ and resorts to the L1), which is then negotiated with Learner 2 until a decision is made.

In this study, self-repairs are counted as part of LRE but are also reported separately.³ For self-repairs, we adopt the coding system by Gilabert (2007c) based on Kormos (1998). In the calculation both error and non-error repairs are coded. While error repairs are the consequence of faulty morphology, syntax, word choice, or phonology, non-error repairs include what Kormos (1998) has labeled as different and appropriateness repairs. The main criterion differentiating between error and non-error repairs was that in the former the learner repaired errors in morphology, syntax, morphology, or phonology, while in the latter repair was made either to achieve precision, resolve ambiguity, or provide more appropriate information (i.e., different and appropriateness repairs in Kormos' taxonomy). For example:

Learner: OK so I have only two pictures which is open because I think I think in the story the man goes upstairs to see what happen what happened what is happening (ERROR REP) upstairs

Learner: you will see a building uh a red building (NON-ERROR REP)

2.3.2. *Statistical instruments, analyses, transcription and coding.* Descriptive statistics providing information about means and standard deviations were used; repeated-measures analyses of variance (ANOVA) were employed for the calculation of main effects, and pairwise comparisons to identify the exact location of differences in the case of affective variables, which were normally distributed; since none of the dependent variables was normally distributed, non-parametric Friedman repeated measure tests were used followed by pairwise Wilcoxon Signed Ranked tests. Significance levels were set at $\alpha = .05$.

The CA mode of CHILDES (MacWhinney 1995) was used for the transcription of the 162 tasks. Both intrarater and interrater calculations were applied in the transcription and coding of the narratives. The transcription of the narratives was carried out by the three researchers. Intrarater reliability reached 95 %, and interrater⁴ agreement out of a randomly selected sample of 10 % percent of the data reached 90 %.

3. There are two reasons for this, firstly, when calculating LREs during the pilot study (Gilabert 2007a) only self-repairs clearly triggered a high number of instances with very few episodes of other types of LREs; secondly, Task Complexity has been shown to have a strong effect on self-repair behavior (Gilabert 2007c), and therefore it was thought interesting to report self-repairs separately for the sake of comparability.

4. Interrater reliability was calculated by means of the percentage of agreement.

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	∂^2
Difficulty	5	164.660	10.812	.000 ^b	.190
Stress	5	46.302	3.484	.005 ^a	.070
Confidence	5	37.927	3.934	.002 ^a	.079
Interest	5	3.039	.486	.787	.010
Motivation	5	15.027	2.047	.073	.043

Df Degrees of freedom

∂^2 partial eta squared (effect size)

a $p < .05$

b $p < .001$

3. Results

3.1. *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 while no significant main effects were found for interest and motivation. Learners graded the three complex versions of the tasks as more difficult, and pairwise comparisons showed that learners perceived the complex version of tasks to be significantly more difficult for the narrative task ($p < .001$), instruction-giving map task ($p < .05$) and the decision-making task ($p < .05$).⁵ Pairwise comparisons for stress showed differences to exist between the simple and complex versions of the narrative task ($p < .001$) but not between the simple and complex versions of the map task or the decision-making task. As for confidence, results show that learners felt significantly less confident when performing the complex narrative, with no differences existing between simple and complex versions of the map task or the decision-making

5. The tables show the raw number of conversational moves per task. However, differences may not be representative due to differences in text length. That is why the number of tokens used for each task was compared between the simple and complex versions of tasks. Only the instruction-giving map task displayed any significant differences. Hence a compensatory measure was used as in Gilabert (2007c) in which each of the measures was divided by the total number of words. Confirmation checks displayed a strong trend in the same direction as raw number ($p = .09$) and comprehension checks showed a significant impact ($p < .05$) of task complexity on interaction. Again, this was also true for language-related episodes ($p < .05$) but not for repairs or recasts. In sum, the measure used to compensate for differences in text length showed a similar impact of Task Complexity on interaction as the measure of raw numbers except in the case of repairs.

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
Clarification requests	.19	.62	.63	1	3.15	2.68	2.81	2.52	.85	1.34	.67	0.96
Confirmation checks	.12	.31	.37	.62	.63	1.33	1.62	3.49	.63	1	.67	.78
Comprehension checks	.22	.42	.00	.00	1.41	2.06	2.31	2.72	.00	.00	.00	.02
Recasts	.07	.26	.15	.36	.37	.74	.19	.48	.04	.97	.37	.74
Language-related episodes	5.07	4.47	7.07	4.38	5.88	3.14	8.38	5.82	5.66	5.33	6.40	5.16
Repairs	3.85	3.67	5.93	3.74	4.04	2.29	6.35	5.42	3.96	2.95	5.19	4.35

Table 3. *Friedman repeated measures results*

Dependent variable	χ^2	<i>p</i> -value
Clarification requests	61.454	.000 ^b
Confirmation checks	17.232	.004 ^a
Comprehension checks	54.270	.000 ^b
Recasts	5.369	.372
Language-related episodes	15.481	.008 ^a
Repairs	14.331	.014 ^a

a $p < .05$ b $p < .001$

task. Neither interest nor motivation showed any significant differences between the simple and complex versions. Only motivation shows a strong trend ($p = .07$) in the case of the map task, which may suggest that, had the task been slightly more complex, learners would have found it less motivating. It can therefore be concluded that, overall, more complex tasks were perceived as more difficult without the learners finding them less interesting or motivating.⁶ Finally, the sequence of presentation of the tasks did not affect learners' perception as shown by the results of repeated-measures ANOVAs of the five affective variables in which "set" was included as a between-subject factor.

3.2. *Results of Hypothesis 1: Interactional moves as affected by task complexity*

None of the measures showed any interaction between sequence of presentation and Task Complexity to be significant. As seen in Table 3, main effects were significant for all measures except for recasts.

Hypothesis 1 was largely confirmed but with different results being obtained for each task. As shown in Table 4, the complex narrative triggered significantly greater numbers of clarification requests ($p < .05$) and a strong trend ($p = .070$) in the case of confirmation checks. A significantly lower number of comprehensions checks ($p < .05$), however, was obtained in the complex version of the task.⁷ Task complexity also had a significant impact on language-

6. These findings for learner perceptions of complex tasks are in line with previous studies using the same metric (e.g., Ishikawa 2007; Robinson 2001a, 2007b; Gilabert 2007b, 2007c).

7. A closer look at the data revealed that only 6 out of the 27 dyads in the study used 1 comprehension check each in the simple version, while none were used by any of dyads in the complex version. This explains the significant differences between the two versions of the task, but it also presents a distorted image given the low number of occurrences of such episodes

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

Dependent variable	Narrative		Map		Fire chief	
	Z	p-value	Z	p-value	Z	p-value
Clarification requests	-2.279	.023 ^a	-.764	.445	-.854	.393
Confirmation checks	-1.809	.070	-2.309	.021 ^a	-.179	.858
Comprehension checks	-2.449	.014 ^a	-2.032	.042 ^a	-.414	.157
Recasts	-1.000	.317	-.845	.398	-.105	.917
Language-related episodes	-2.151	.031 ^a	-2.250	.024 ^a	-1.434	.152
Repairs	-2.293	.022 ^a	-2.509	.012 ^a	-2.468	.014 ^a

a $p < .05$ b $p < .001$

related episodes ($p < .05$) and the number of repairs ($p < .05$), but not on recasts. The results of the narrative task therefore largely confirmed Hypothesis 1. As for the map task, confirmation checks ($p < .05$) and comprehension checks ($p < .05$) showed a significant impact of task complexity on interaction. Noteworthy is the fact that our results show comprehension checks to be exclusively associated with the map task but not with either of the other two tasks. Again, this was also true for language related episodes ($p < .05$) and repairs ($p < .05$), and not for recasts. Finally, the decision-making task showed no impact of task complexity on any of the measures except for the number of repairs ($p < .05$). In summary, we can say that task complexity had a strong impact on most measures of interaction except for recasts, and only in the case of the narrative reconstruction task and the instruction-giving task, but not in the case of the decision-making “fire chief task”. The number of repairs was always significantly higher when the three task types were performed under complex conditions.

3.3. *Results of Hypothesis 2: Performance among task types*

In general, as an answer to the second question in this study, the instruction-giving map task is the one that generated the highest number of clarification requests and comprehension checks. Under simple conditions, the narrative reconstruction task triggered the lowest number of negotiation of meaning episodes, particularly in the case of clarification requests and comprehension

in this task. The interpretation here, then, is that comprehension checks were not significantly associated with the narrative task.

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

	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
Clarification requests	.000 ^b	.009 ^a	.000 ^b	.000 ^b	.873	.000 ^b
Confirmation checks	.064	.017 ^a	1	.017 ^a	.087	.187
Comprehension checks	.006 ^a	.014 ^a	.001 ^a	.000 ^b	.157	.001 ^a
Recasts	.070	.107	.963	.608	.084	.437
Language-related episodes	.204	.497	.328	.327	.458	.103
Repairs	.456	.892	.644	.922	.234	.257

a $p < .05$ b $p < .001$

checks, which were significantly lower than those of the instruction-giving map task and the decision-making fire chief task. It is interesting to note that the differences in terms of negotiation of meaning between the narrative task and the decision-making task disappeared under complex performance, both generating similar numbers of moves. Neither recasts, language-related episodes nor repairs showed any significant differences among the different task types regardless of the conditions under which they were performed. We can therefore conclude that Hypothesis 2 was only partially confirmed.

4. Discussion and conclusions

4.1. Hypothesis 1

The results in this study suggest that manipulating task complexity along resource-directing variables such as the degree of displaced past time reference or the number of elements included in that task has the potential to generate higher numbers of interactional moves. This, however, is true for some tasks but not for others. Except for recasts, the narrative task displayed either significant differences or a strong trend between the simple and complex versions, with the complex task generating more interaction. This is consistent with the findings of Nuevo (2006), who found the unstructured story retelling to generate more confirmation checks, comprehension checks, and repairs. We believe that the prompt to reconstruct the story in there-and-then forced learners to stretch their attentional and memory resources, to hold the vignettes in memory for their scrutiny, and to be precise about the information they needed to communicate. As a consequence, a higher number of clarification demands and confirmation checks were needed to work out the plot of the story. Additionally, as was reported in Gilabert (2007c) for the same task in a monologic mode, the higher number of self-repairs provides evidence of enhanced attention to lexical and grammatical encoding during task performance. Similarly, the map task generated more confirmation checks and comprehension checks in the complex version. Navigating along three different axes and referring to similar elements imposed a certain degree of precision in learners' description of the map if they wanted their directions to be followed accurately. As suggested by the Cognition Hypothesis, task demands requiring the conceptualization of and reference to more complex spatial concepts may have prompted more revision of comprehension and production. This may explain why listeners kept confirming whether their construal matched the speaker's descriptions, and speakers repeatedly asked whether they were being understood. Furthermore, in the complex version of the map task, as also reported by Gilabert (2007c), more attention was devoted to the morphosyntactic and

lexical features of their discourse. The decision-making task showed the most deviant behavior of all tasks. While it is true that more attention was paid to message encoding, as shown by the higher number of self-repairs, no other measures showed any significant differences between the simple and complex versions. We would speculate that this was due to the open nature of the task. In the decision-making task no single solution was required, and therefore all learners needed to do was to agree on one of the many possible solutions. Once learners shared the information that was different in the two buildings, which may have generated part of the interaction reported in the results, their goal was to reach an agreement on what action to take and their performative sequence. The kind of discourse generated by such a task may not be captured by clarification requests, comprehension checks, and confirmation requests. However, a qualitative look at the data pointed out the use of many other pragmatic moves such as suggestions, expressions of agreement and disagreement, as well as other pragmatic moves which were not captured by the measures employed here. Future studies could investigate whether task complexity affected interaction beyond the moves analyzed in this study. The low incidence of recasts across all tasks also deserves an explanation. The number of recasts produced by the learners in our study does not differ considerably from the ones reported in Gass et al. (2005). There are at least two possible explanations as to why students do not recast their interlocutor's production. The first reason may be their lack of training in recasting in an EFL learning context that still fosters traditional grammar learning with little focus on communication and interaction. Secondly, there may exist a cultural component in which learners do not feel they should be recasting their conversational partners' errors. In sum, increasing task complexity, through interactional episodes and through self-repair, led learners to reorganize their messages, to be more precise, and to push their production to achieve the successful completion of the task.

Finally, our findings with regard to the impact of complexity on interaction cast a doubt on models of attention that claim too much focus on meaning deviates attention from form. Students in the three tasks certainly focused on meaning to work out the different solutions but were still able to negotiate meaning (through clarification requests, confirmation checks, and comprehension checks) and to devote some attention to form as seen by the results of language-related episodes and repairs. Multiple-resource models such as the one advanced by Wickens (1989, 1992, 2007) would therefore seem to be more appropriate in explaining attention switches between meaning and form than single resource, limited capacity accounts of attention.

4.2. *Hypothesis 2*

In the results section we saw that there were differences in the associations between certain task types and certain interactional moves. While most conversational episodes were used in the three tasks, the instruction-giving map task showed the highest number of instances of clarification requests and comprehension checks. This may point towards the fact that it is probably the task that requires the highest precision in the transfer of information in order to locate the path and the landmarks along the path. While in the narrative task, and clearly in the decision-making task, there was room for speculation about the sequence of events, in the map task only one solution was possible in each case, which may have raised the need to clarify and confirm whether information had been transmitted properly. This is also consistent with previous findings which have looked at the differential impact of closed and open tasks on dialogic production (Long 1990; Crookes and Rulon 1985). We also saw that the differences between the narrative task, which generated significantly less negotiation of meaning in the simple version as compared to the decision-making task, and the decision-making task disappeared in the complex version. This also confirms the impact of Task Complexity on interaction.

4.3. *Conclusions, limitations, and future research*

In our view, the results in this study provide confirmatory evidence for both the Interaction Hypothesis and the Cognition Hypothesis. If we believe the arguments advanced by interactionist researchers and, more importantly, the evidence that has already started showing clear connections between interaction and acquisition, we may conclude that increasing the cognitive complexity of tasks in this study may contribute to enhance interaction which, in turn, has the potential to help learners in their interlanguage development (Gass and Mackey 2007). However, we have also seen that the predictions of both theoretical frameworks may not quite hold for all task types. This certainly emphasizes the need to further explore the internal features of tasks and tap into the nature of the discourse that they generate. Two limitations that need to be pointed out may be the number of participants and their proficiency level. Firstly, the small number of learners participating in the study may not make results generalizable enough. Secondly, the study used subjects with similar levels of proficiency (at an upper-intermediate level), who may behave differently from learners at lower levels. In future studies, measuring individual differences other than proficiency would certainly provide better explanations for the impact of Task Complexity on interaction, especially those tapping into attention and memory. Additionally, systematically combining the variables

operationalized in this study with other variables such as \pm planning time would more closely approximate performance conditions to those of real-life target tasks, and would provide a more complete picture of how cognition and interaction affect production and learning.

University of Barcelona
<rogergilabert@ub.edu>
<juliabaron@ub.edu>
<allanes@ub.edu>

Appendix

Simple. Narrative reconstruction task

Student A

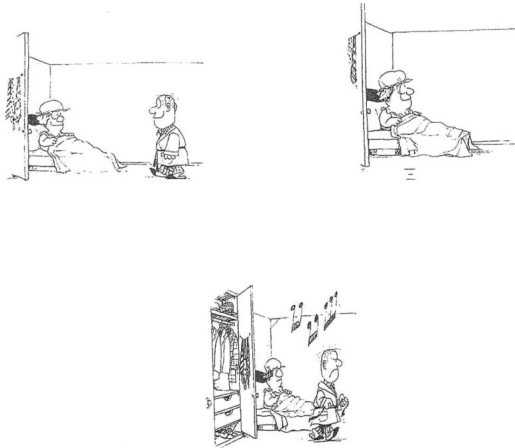


Student B

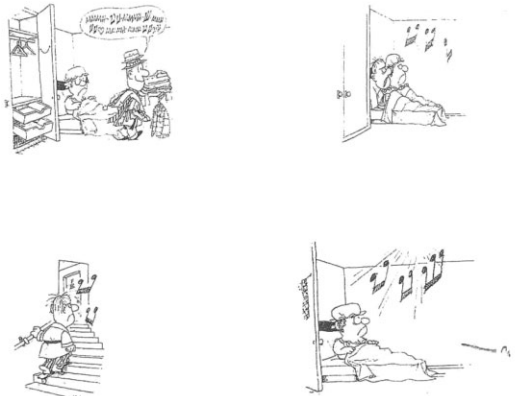


Complex. Narrative reconstruction task

Student A

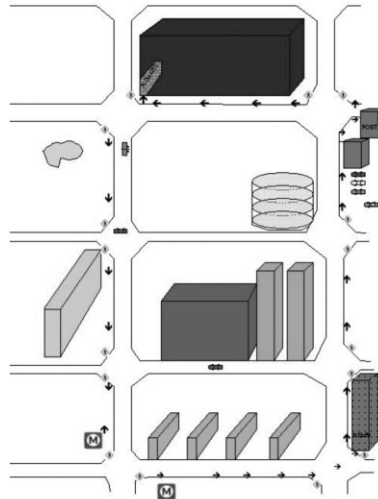


Student B

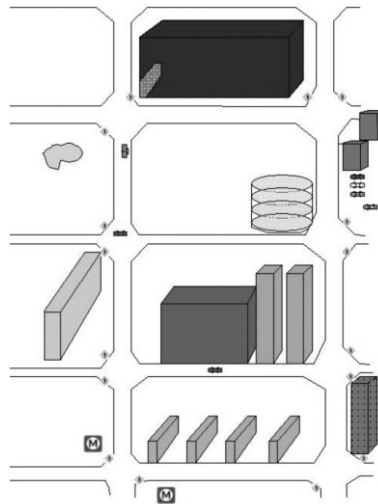


Simple. Instruction-giving map task

Instruction-giver

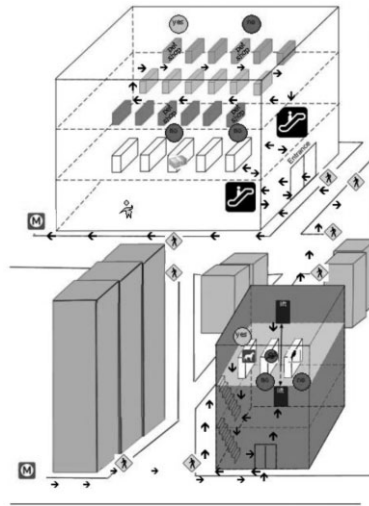


Instruction-receiver

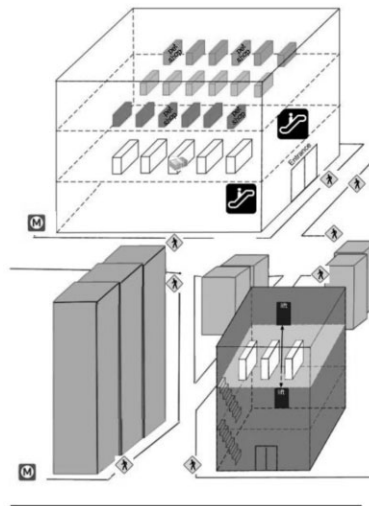


Complex. Instruction-giving map task

Instruction-giver

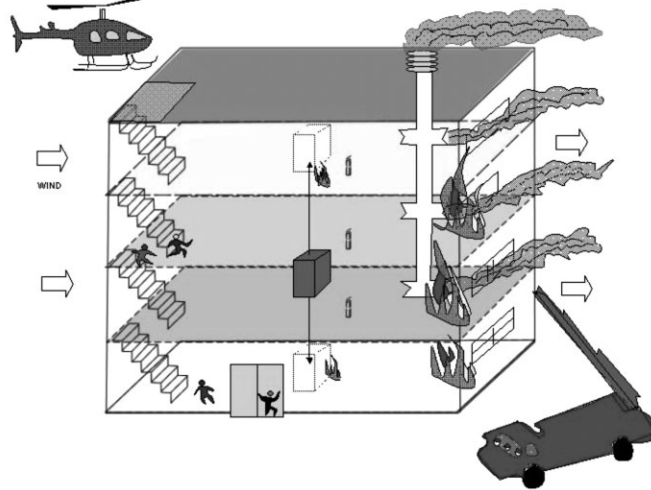


Instruction-receiver

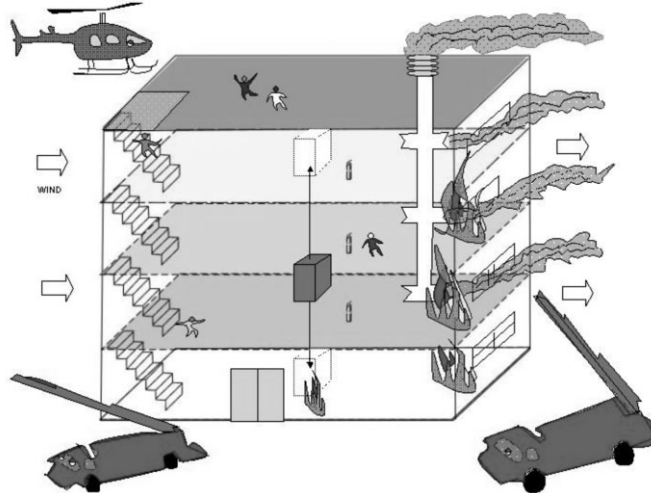


Simple. Decision-making Fire Chief task

Prompt for Student A

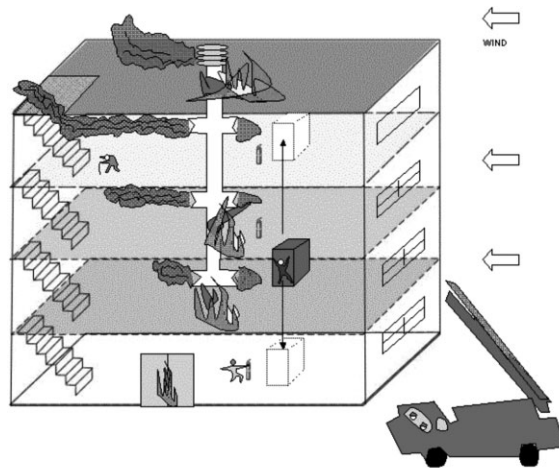


Prompt for Student B

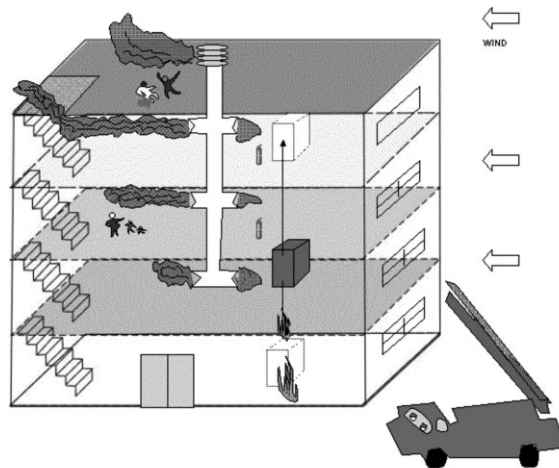


Complex. Decision-making Fire Chief task

Student A



Student B



References

- Alegría de la Colina, Ana, and Maria del Pilar García Mayo (2007). Attention to form across collaborative tasks by low proficiency learners in an EFL setting. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García-Mayo (ed.), 91–116. Clevedon: Multilingual Matters.
- Allport, Alan (1987). Selection for action: Some behavioral and neurophysiological consequences of attention and action. In *Perspectives on Perception and Action*, Herbert Heuer and Andries F. Sanders (eds.), 395–419. Hillsdale, NJ: Erlbaum.
- 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.
- Crookes, Graham, and Kathryn A. Rulon (1985). Incorporation of corrective feedback in native speaker/non-native speaker conversation. In *Technical Report No. 3*, The Center for Second Language Classroom Research, Social Science Research Institute, University of Hawai'i at Manoa.
- Gass, Susan (1997). *Input, Interaction, and the Second Language Learner*. Mahwah, NJ: Lawrence Erlbaum.
- Gass, Susan (2003). Input and interaction. In *Handbook of Second Language Acquisition*, Catherine Doughty and Michael H. Long (eds.), 224–255. Oxford: Blackwell.
- Gass, Susan M., and Allison Mackey (2007). Input, interaction and output: An overview. *AILA Review* 19: 3–17.
- Gass, Susan, Allison Mackey, and Lauren Ross-Feldman (2005). Task-based interactions in classroom and laboratory settings. *Language Learning* 55 (4): 575–611.
- Gilabert, Roger (2005). Task complexity and L2 narrative oral production. Unpublished doctoral dissertation, University of Barcelona.
- Gilabert, Roger (2007a). Effects of manipulating task cognitive complexity on interactional modifications during L2 oral performance. Paper presented at the European Second Language Association (EUROSLA) 17 Conference, University of Newcastle, 11–14 September.
- Gilabert, Roger (2007b). The simultaneous manipulation of task complexity along planning and +/- Here-and-Now: Effects on L2 oral production. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García Mayo (ed.), 44–68. Clevedon: Multilingual Matters.
- Gilabert, Roger (2007c). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics* 45: 215–240.
- Gopher, Daniel (1992). Analysis and measurement of mental workload. In *International Perspectives on Psychological Science: Vol. 2, State of the Art*. Géry d'Ydewalle, Paul Eelen, and Paul Bertelson (eds.), 265–291. Hillsdale, NJ: Lawrence Erlbaum.
- Han, ZhaoHong (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly* 36: 543–572.
- Ishida, Midori (2004). Effects of recasts on the acquisition of the aspectual form -te i-(ru) by learners of Japanese as a foreign language. *Language Learning* 54: 311–394.
- Iwashita, Noriko (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition* 25: 1–36.
- 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.
- Lee, Eliza C., and Leslie Rescorla (2002). The use of psychological state terms by late talkers at age 3. *Applied Psycholinguistics* 23: 623–641.
- Long, Michael H. (1983). Native speaker/non-native speaker conversation in the second language classroom. In *On TESOL '82: Pacific Perspectives on Language Learning and Teaching*, Mark A. Clarke and Jean Handscombe (eds.), 207–225. Washington, DC: TESOL.

- Long, Michael H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly* 24: 649–666.
- Long, Michael H. (1996). The role of the linguistic environment in second language acquisition. In *Handbook of Language Acquisition. Vol. 2: Second Language Acquisition*, William C. Ritchie and Tei K. Bhatia (eds.), 413–468. New York: Academic Press.
- Lyster, Roy (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition* 20: 51–81.
- Lyster, Roy, and Leila Ranta (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition* 19: 37–66.
- Mackey, Alison, and Jaemyung Goo (2007). Interaction research in SLA: A meta-analysis and research synthesis. In *Conversational Interaction in Second Language Acquisition*, Alison Mackey (ed.), 407–452. Oxford: Oxford University Press.
- Mackey, Alison, and James Philp (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal* 82: 338–356.
- MacWhinney, Brian (1995). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Lawrence Erlbaum.
- McDonough, Kim (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition* 27: 79–103.
- McDonough, Kim (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In *Conversational Interaction in Second Language Acquisition*, Alison Mackey (ed.), 323–338. Oxford: Oxford University Press.
- McDonough, Kim, and Alison Mackey (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning* 56: 693–720.
- Meara, Paul, and James Milton (2003). *X_Lex: The Swansea Vocabulary Levels Test*. Newbury: Express.
- Michel, Marije, Folkert Kuiken, and Ineke Vedder (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics* 45: 241–259.
- Nassaji, Hossein (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning* 57: 511–548.
- Nichols, Howard, Patsy M. Lightbown, and Nina Spada (2001). Recasts as feedback to language learners. *Language Learning* 51: 719–758.
- Nuevo, Ana-Maria (2006). Task complexity and interaction. Unpublished doctoral dissertation, Georgetown University.
- Pica, Teresa (1994). Research on negotiation: What does it reveal about second language learning, conditions, processes, outcomes? *Language Learning* 44: 493–527.
- Pica, Teresa, Ruth Kanagy, and Joseph Falodun (1993). Choosing and using communication tasks for second language instruction. In *Tasks and Language Learning. Integrating Theory and Practice*, Graham Crookes and Susan Gass (eds.), 9–34. Clevedon: Multilingual Matters.
- Quesada, José F., Walter Kintsch, and Emilio Gómez-Milán (2001). Theory of complex problem solving using the vector space model (part 2). Latent Semantic Analysis applied to empirical results from adaptation experiments. In *Cognitive Research with Microworlds*, José J. Cañas (ed.), 147–158. Taylor and Francis.
- Révész, Andrea (2007). Focus on form in task-based language teaching: Recasts, task complexity, and L2 learning. Unpublished doctoral dissertation, Columbia University.
- Robinson, Peter (1995). Task complexity and second language narrative discourse. *Language Learning* 45: 99–140.
- Robinson, Peter (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22 (1): 27–57.
- Robinson, Peter (2001b). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In *Cognition and Second Language Instruction*, Peter Robinson (ed.), 287–318. Cambridge: Cambridge University Press.

- Robinson, Peter (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies* 21 (2): 45–105.
- Robinson, Peter (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics* 55: 1–32.
- Robinson, Peter (2007a). Criteria for classifying and sequencing pedagogic tasks. In *Investigating Tasks in Formal Language Learning*, Maria del Pilar García Mayo (ed.), 7–26. Clevedon: Multilingual Matters.
- Robinson, Peter (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics* 45: 193–213.
- Robinson, Peter, and Roger Gilabert (2007). Task complexity, the cognition hypothesis and second language learning performance. *International Review of Applied Linguistics* 45: 161–176.
- Schmidt, Richard (1990). The role of consciousness in second language learning. *Applied Linguistics* 11: 129–158.
- Shatz, Marilyn, Henry M. Wellman, and Sharon Silber (1983). The acquisition of mental verbs: a systematic investigation of the first reference to mental state. *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.
- Skehan, Peter, and Pauline Foster (2001). Cognition and tasks. In *Cognition and Second Language Instruction*, Peter Robinson (ed.), 183–205. Cambridge: Cambridge University Press.
- Swain, Merrill (1995). Three functions of output in second language learning. In *Principles and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*, Guy Cook and Barbara Seidelhofer (eds.), 125–144. Oxford: Oxford University Press.
- Swain, Merrill (1998). Focus on form through conscious reflection. In *Focus on Form in Classroom SLA*, Catherine Doughty and Jessica Williams (eds.), 64–81. New York: Cambridge University Press.
- Swain, Merrill (2005). The output hypothesis: Theory and research. In *Handbook of Research in Second Language Teaching and Learning*, E. Hinkel (ed.), 471–483. Mahwah, NJ: Lawrence Erlbaum.
- Swain, Merrill, and Sharon Lapkin (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics* 16: 370–391.
- Swain, Merrill, and Sharon Lapkin (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal* 82: 320–337.
- Swain, Merrill, and Sharon Lapkin (2001). Focus on form though collaborative dialogue: Exploring task effects. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*, Martin Bygate, Peter Skehan and Merrill Swain (eds.), 99–118. Harlow: Longman.
- Wickens, Christopher D. (1989). Attention and skilled performance. In *Humans Skills*, D. H. Holding (ed.), 71–105. New York: John Wiley.
- Wickens, Christopher D. (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins.
- Wickens, Christopher D. (2007). Attention to the second language. *International Review of Applied Linguistics* 45: 177–191.