Studies in Second Language Acquisition, 2013, **35**, 689–725.
doi:10.1017/S0272263113000429

THE IMPACT OF COGNITIVE COMPLEXITY ON FEEDBACK EFFICACY DURING ONLINE VERSUS FACE-TO-FACE INTERACTIVE TASKS

Melissa Baralt
Florida International University

Informed by the cognition hypothesis (Robinson, 2011), recent studies indicate that more cognitively complex tasks can result in better incorporation of feedback during interaction and, as a consequence, more learning. It is not known, however, how task complexity and feedback work together in computerized environments. The present study addressed this gap by investigating how cognitive complexity in face-to-face (FTF) versus computer-mediated communication (CMC) environments mediates the efficacy of recasts in promoting second language development. Eighty-four adult learners of Spanish as a foreign language at a mid-Atlantic university were randomly assigned to a control group or one of four experimental groups. The experimental groups engaged in one-on-one interaction and received recasts on the Spanish past subjunctive but differed according to (a) whether or not they had to reflect on another person's intentional reasons during the task and (b) whether they interacted in FTF or CMC environments. Learning was measured with two production

The data reported on here are from my Ph.D. dissertation completed at Georgetown University. I would like to thank my mentor, Ron Leow, for his guidance and encouragement. I am also incredibly grateful for the support of my dissertation committee: Peter Robinson (who always told me to read outside the field), Andrea Révész, and Rusan Chen. Special thanks go to Laura Gurzynski-Weiss, Rebecca Sachs, and Julio Torres, who helped me brainstorm ways to elicit the targeted linguistic item in conversational interaction. I am humbled by their counsel and friendship. Last but not least, I am grateful for my students. Watching them acquire and use another language is the most fascinating privilege and was the source of inspiration for this study. Many thanks are due to the Department of Spanish and Portuguese at Georgetown, to CNDLS, and to the anonymous reviewers who provided helpful comments on this manuscript. Any remaining errors are my own.

Correspondence concerning this article should be addressed to Melissa Baralt, Department of Modern Languages, Florida International University, Modesto A. Maidique Campus, Deuxième Maison (DM) 499, Miami, FL 33199. E-mail: mbaralt@fiu.edu

© Cambridge University Press 2013

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

tasks and a multiple-choice receptive test in a Pretest-Posttest 1-Posttest 2 design. Results revealed that in the FTF mode, performing the cognitively complex task while receiving recasts led to the most learning. In the CMC mode, the cognitively complex task + recasts was not effective. Instead, the cognitively simple task led to the most development in CMC. The study also found that judgments of time on task were the only independent measure of cognitive complexity that held across mode.

Two guiding methodological principles of task-based language teaching are that engaging in conversational interaction is an effective way to learn a foreign language and that, at times, it will be necessary to draw learners' attention to linguistic features in the context of meaningful communication to facilitate learning. How to best achieve a directing of attention to form during interaction is currently an area of prolific research in SLA, with different types of focus-on-form techniques—and, in particular, the recast—being heavily investigated (Doughty & Long, 2003; Doughty & Williams, 1998; Long, 2000; Long & Robinson, 1998; see also Lyster & Ranta, 1997). A recast is a reformulation of a learner's utterance during which the erroneous element is corrected but the focus stays on meaning (see R. Ellis, 2003; Long, 2007; Mackey, 2007; Nicholas, Lightbown, & Spada, 2001; Samuda & Bygate, 2008, for reviews). The theoretical justification for recasts is that, because they rephrase the learner's original utterance, the learner will (a) already have comprehension of the message, (b) have more freed attentional resources to allocate, and (c) be more motivated and attentive to the input; all of these are factors that should facilitate noticing of the corrected form (Long, 2007).

The effectiveness of recasts is centered on the assumption that learners notice them and their corrective function. Research has shown, however, that the benefits that come from noticing recasts are constrained by various learner-internal variables (e.g., Goo, 2012; Philp, 2003; Révész, 2011; Trofimovich, Ammar, & Gatbonton, 2007). Catering to these ability factors is one of the greatest challenges in the classroom setting. What teachers and researchers do have control over, however, is the design of tasks used to engage learners in interaction. It may be possible that a task's engineering can differentially focus learners' attention to features of the input, thereby mediating the efficacy of recasts. Informed by the cognition hypothesis of task-based language learning (Robinson, 2001a, 2001b, 2007, 2010, 2011; Robinson & Gilabert, 2007), an incipient body of research in this area is investigating whether or not certain task design features—namely, those related to the cognitive complexity of the task—have the capacity to induce learners to notice recasts better than other

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

task features. According to the cognition hypothesis, task complexity, or the "attentional, memory, reasoning, and other information processing demands" that are required of the language learner as a result of a task's design (Robinson, 2001b, p. 29), is hypothesized to promote better form-meaning mappings during task-based interaction. Its claim is that increases in the cognitive complexity of a task will push learners to recognize those linguistic capacities they do not have, and they will thus be set up to better notice and incorporate feedback in the input so that they can successfully carry out the task. So far, a few studies (e.g., Révész, 2009) have shown that recasts alongside more cognitively complex tasks may work best to promote language development.

But does this combined benefit of task complexity and feedback hold for the computerized environment? There has been a recent increase in the development of long-distance language programs, and, as such, there is a need to examine how the online environment mediates task design effects. So far, no research exists that compares task-based interaction and feedback in the face-to-face (FTF) mode with the computer-mediated communication (CMC) mode. The present study aims to fill this gap by bringing together the lines of investigation on task complexity, feedback, and FTF versus CMC environments. Testing the cognition hypothesis, this study examines how conversational interaction in two different environments, FTF and CMC, affects learning when students receive recasts during tasks of different levels of cognitive complexity.

THEORETICAL BACKGROUND

The Cognition Hypothesis of Task-Based Language Learning

The cognition hypothesis of task-based language learning (Robinson 2001a, 2001b, 2007, 2010, 2011; Robinson & Gilabert, 2007) advocates that pedagogic tasks be presented to learners in an order of gradually increasing complexity. Performing tasks that gradually increase in complexity will lead to increased automatization and eventually—on the most complex version—a "restructuring of the current interlanguage system" (Robinson, 2010, p. 244). When designing tasks for language learning, Robinson (2010) distinguishes between resource-dispersing variables and resource-directing variables. *Resource-dispersing* variables place performative demands on the learner, such as having to perform a task without time to preplan one's linguistic output. Increased complexity along resource-dispersing features helps to improve automatization, which involves faster and quicker retrieval of resources from memory. *Resource-directing* variables place cognitive demands on the learner, such as requiring reasoning alongside the relaying of a past event. Increases in the complexity of these task variables are argued by Robinson

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

to foster interlanguage development, because having to communicate more complex cognitive ideas is synonymous with needing more complex syntactic resources. When a task is more complex along resource-directing dimensions, Robinson hypothesizes that learners are pushed to recognize features of the language that they do not have; the only way they can meet the demands of the task is by employing specific linguistic features. This may prompt the learner to be more tuned to and receptive of feedback that addresses those features so that the task can be carried out.

Two predictions of the cognition hypothesis were tested in this study:

1. If a task's level of cognitive complexity is increased, this will "result in greater attention to, and uptake of, forms made salient during the provision of *reactive* Focus on Forms techniques such as recasts" (Robinson, 2011, p. 18).
2. All learners will perceive cognitively complex tasks as more difficult and more stressful than cognitively simple tasks (Robinson, 2007).

The Cognition Hypothesis and Conversational Interaction

In these predictions, the cognition hypothesis makes reference to "attention," "uptake," "forms made salient," and "Focus on Forms techniques" (Robinson, 2011, p. 18). All of these are key psycholinguistic constructs that both theoretical and empirical work have used to explain how conversational interaction works. Features of interaction that are posited to facilitate learning include (a) the negotiation for meaning work in which learners and their interlocutors engage, (b) feedback as reactive responses to erroneous production, and (c) opportunities for learners to modify their output to make it more comprehensible. Feedback may be one of the most important aspects of interaction because it informs the learner that his or her output was nontargetlike and can encourage the learner to make a comparison between the erroneous and correct forms, which can lead to development. This crucial comparison of forms during interaction to promote second language (L2) learning is, of course, contingent on the learner noticing that a correction was made. According to Schmidt (1990), learners must consciously notice forms and the meanings they convey; this noticing is "the necessary and sufficient condition for converting input to intake" (p. 129). Learners' responses to feedback, such as the production of uptake, are argued to be developmentally useful in L2 acquisition because they may promote learners' awareness of the corrected form as being different than their nontargetlike form.

Considering these cognitive underpinnings of conversational interaction, it is arguable that task complexity can be a cause for interaction-driven learning. For one, and as stipulated by Nuevo (2006), "more complex tasks may elicit more linguistically complex input [and output] that

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

may not be comprehensible to [the] interlocutor" (p. 70) and thus may lead to the interactive negotiation work theorized to facilitate L2 acquisition. Second, it may very well be the case that "task demands are a powerful determinant of what is noticed" (Schmidt, 1990, p. 143) and can modulate what is learned because, according to the noticing hypothesis, what is learned must also be noticed (Schmidt, 1990). In conjecturing how input is noticed, Schmidt (1990) has argued that "how the task forces the material to be processed" determines whether or not learning will result (p. 143). If more complex tasks have a higher noticing effect, as Robinson predicts, cognitive complexity could be a task design variable that induces greater noticing of feedback.

Studies Operationalizing the Cognition Hypothesis

Studies investigating Robinson's cognition hypothesis have fallen into three main areas. The first has examined the effects of cognitive complexity on one- or two-way oral or written production and has looked at dependent variables such as linguistic complexity, accuracy, and fluency (e.g., Ishikawa, 2007; Kuiken & Vedder, 2007; Michel, 2011; Robinson, 2001b). The second has explored ways in which task complexity affects features of two-way or group interaction (e.g., Gilabert, Barón, & Llanes, 2009; Kim, 2009; Nuevo, 2006; Révész, 2011; Robinson, 2001b). The third strand of research has explored L2 development as a result of performing more or less cognitively complex tasks along resource-directing variables (Kim, 2012; Kim & Tracy-Ventura, 2011; Nuevo, 2006; Révész, 2009). For the purpose of the present investigation, this third strand will be reviewed here. Each of the aforementioned four studies examined L2 development as a result of carrying out an interactive task that was more or less complex. Nuevo (2006), Kim (2012), and Kim and Tracy-Ventura (2011) explored naturally occurring peer feedback or language-related episodes (LREs),[1] whereas Révész (2009) examined task complexity in conjunction with reactive feedback provided by an expert interlocutor.

In Nuevo's (2006) study, learners of English as a L2 were paired with other learners and carried out a narrative and a decision-making task. Nuevo operationalized cognitive complexity with the resource-directing variable +/– causal reasoning and examined how task complexity led to different features of interaction (i.e., learners' use of confirmation checks, recasts, clarification requests, etc.) as well as resultant L2 development. The tasks naturally elicited English locative prepositions and the past tense. Learning was measured with an oral production task and a grammaticality judgment test. Nuevo's results indicated no differences in learning between the two groups and, therefore, did not lend support to predictions of the cognition hypothesis.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Kim (2012) also investigated interaction between dyadic pairs but did so within classroom contexts over the course of a semester. Cognitive complexity was operationalized as +/– reasoning demands and +/– few elements and was designed on a continuous scale that ranged from *simple* to +*complex* and ++*complex*. Four intact English as a foreign language (EFL) classes participated in the study, three acting as the experimental groups and one class as the comparison control group. Individual and paired oral production tasks as well as a written metalinguistic test were used to measure learning. Kim found that the class that always performed the ++complex tasks produced the most LREs for English question formation and also achieved the greatest advancement in the development of this structure. In contrast to Nuevo (2006), Kim's study did provide support for the cognition hypothesis.

As a follow-up study, Kim and Tracy-Ventura (2011) examined task complexity and how it, alongside anxiety as a learner individual difference variable, mediated task-based interaction in the classroom. Using the same task complexity variables as Kim (2012), the researchers randomly assigned participants to one of three groups (i.e., simple, +complex, and ++complex), and task-based interaction was carried out and recorded in three intact EFL classrooms over a period of 2 weeks. They found that the ++complex group achieved the highest gains in the development of the English past tense on the posttests, followed by the +complex group and then the simple group. The researchers also reported that low-anxiety learners performed significantly higher than the high-anxiety learners on the delayed posttest.

Révész's (2009) study operationalized +/– contextual support—or the presence or lack of a photo while having to describe the scene—as a means of increased cognitive complexity. Learners interacted one on one with the researcher and performed the simple or the complex task. Half of the learners received recasts on erroneous production of the English past progressive, whereas the other half did not receive feedback. Learning was measured with one written and two oral production tasks. Révész found that the learners who received recasts and did not have contextual support (i.e., the more cognitively complex task) achieved the greatest L2 gains. This was the first study to show that recasts alongside more cognitively complex tasks may lead to greater L2 development.

Measuring Learning as a Result of Interaction with Cognitively Complex Tasks

Three out of the four studies reviewed in the previous section suggest that more complex tasks lead to more learning in that increased task

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

complexity seemed to promote attention to the linguistic forms needed to successfully perform interactive tasks. Arguably, however, results are still inconclusive with regard to whether or not increased complexity positively impacts interaction-driven L2 development. For example, in explaining her results, Nuevo (2006) questioned if her assessments properly captured L2 development. She postulated that the use of tailored items (i.e., items that reflect actual points, reasons, emotions, etc., that learners come up with themselves during the task) on the posttests may best capture interaction-driven learning. When researching L2 development and task complexity, Nuevo called for research designs that use assessments with tailored items to acknowledge what learners contribute to the task. For the present study, a combination of tailored and nontailored items was used on the assessments (see the Materials section). This was done so that learners' own contributions to the interactive task could be taken into account and so that a statistical comparison of groups could be run.

Independent Measures of Task Complexity

When operationalizing the cognition hypothesis, there is another, often overlooked, issue to keep in mind. Norris and Ortega (2009) have pointed out the problematic and circular logic of assuming cognitive complexity as based on dependent measures of linguistic complexity. Cognitive complexity must be measured independently from linguistic production to ensure that the construct has been properly tapped. So far, questionnaires have been the only method used to measure how learners perceive the complexity level of a task (e.g., Gilabert et al., 2009; Kim, 2009; Révész, 2009; Robinson 2001a), with some containing only five items in the instrument, and no study reporting on the instrument's validity. A more robust way to measure complexity may be to record learners' retrospective time estimation by asking them to judge how long they believe it took them to complete the task. Estimating time on task as a measure of cognitive load is supported by literature in the field of psychology, which suggests that the greater the processing demands placed on the learner, the more time he or she will judge has passed (indicating a linear relationship between cognitive demands and estimations of time on task: see, e.g., Block, Hancock, & Zakay, 2010; Fink & Neubauer, 2001; Paas, Tuovinen, Tabbers, & Van Gerven, 2003). The present study employed both questionnaires and retrospective judgments of time on task[2] to measure cognitive complexity.[3]

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Task-Based Interaction in the Computerized Environment

All of the studies reviewed so far on learning outcomes with tasks of varying levels of cognitive complexity have been exclusive to the FTF environment. Since the emergence of online environments, Web 2.0 applications,[4] and even multiuser virtual environments (many of which are now used for distance language education to include online language classes), researchers have started to investigate how learning is differentially experienced in these modes. One online environment that has particularly caught the attention of SLA researchers is CMC chat, which allows for real-time, synchronous conversation via the Internet. The synchronicity of interaction in CMC has been argued to make it similar to oral discourse (Beauvois, 1992; Lamy & Hampel, 2007; Pellettieri, 2000), and yet, the fact that CMC is a hybrid between written and oral conversation renders it a variable of task design that could maximize L2 acquisition. For example, interacting in the CMC mode involves slower turn taking, which is considered to give learners extra time to process input (Beauvois, 1992). This extra time is hypothesized to assist learners in planning their linguistic output while engaging in interaction (Ayoun, 2004; Chapelle, 1998; Sauro & Smith, 2010). In fact, several of the findings from interaction research in the FTF mode have been upheld in the CMC environment, and multiple studies have shown that synchronous CMC is a platform in which learners negotiate for meaning (Blake, 2000; Fernández-García & Martínez-Arbelaiz, 2002; Lee, 2001; Pellettieri, 2000), self-correct (Lai & Zhao, 2006), reformulate their output (Lee, 2001; Salaberry, 2000), and provide and receive feedback (Ayoun, 2004; Lai & Zhao, 2006; Sachs & Suh, 2007; Sauro, 2009). It has also been empirically demonstrated that conversational interaction in the CMC mode can result in L2 development (e.g., de la Fuente, 2003; Sachs & Suh, 2007; Smith, 2005); however, there is still minimal research that explores how different task design features in CMC mediate noticing of feedback and learning. Both Chapelle (2001) and Long (2007) have argued that because feedback given in CMC is visually written out for the learner, this increases the saliency of corrected forms and may thus facilitate the noticing of feedback. Sauro (2009) further explains this characteristic, describing it as an "enduring as opposed to ephemeral record of the interaction" (p. 101). That is, when provided with feedback in FTF interaction, learners are required to cognitively compare corrective recasts with their previous production, which are "memory traces that may have faded" (Sauro, 2009, p. 101). In contrast, comparison of feedback with erroneous production in CMC-based interaction can be done on the computer screen as opposed to in memory. In other words, the command of attentional resources needed to be able to notice recasts in FTF mode (i.e., Doughty, 2001) may be relieved in CMC; this may be why CMC, in and of itself, can be a means to promote attention to form.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

At the same time, interaction in CMC does pose some inherent differences from FTF interaction. Because of the absence of social cues, turn taking in CMC conversation can sometimes be problematic. Lai, Fei, and Roots (2008), for example, reported that the split negotiation routines common to CMC discourse can negatively affect the contingency of recasts in response to errors. Smith (2005) also reported that uptake after focus-on-form episodes has very minimal presence in the CMC mode. If and how these differences affect the utility of interactional feedback in CMC is unknown, and there is no research to date that provides information on how to deal pedagogically with these differences so that the affordances of CMC interaction can be maximized. Notable to this omission is the overall lack of studies on task design for CMC.

Although much of the CMC research base has focused on examining the same interactive features proven to promote L2 acquisition in the FTF environment, there is still little research that measures learning outcomes in CMC, nor is there any research that shows how tasks for interaction-driven learning work differently according to mode. Long (2007) specifically called for research on recasts in CMC to validate their efficacy for online language programs. Despite numerous calls for research—some even more than a decade ago (e.g., Pellettieri, 2000)—on tasks that work best in the computerized environment, research on task design has so far been tangential to CMC. Additionally, it is still an empirical question whether or not transferability of learning from CMC to FTF environments is possible—that is, whether language performance (e.g., complexity, grammaticalization) and learning (e.g., memory of forms provided in the input) achieved by interaction in the CMC mode is transferable to FTF communication. More studies are needed that explore the effects of two-way, interaction-driving learning in both FTF and CMC environments as well as how cognitive complexity is experienced differently in both modes. Not only is this an important next step in discovering which tasks are best at resulting in L2 acquisition, it is also a means to test the cognition hypothesis as a driving force behind the design of tasks in which focus-on-form opportunities are theorized to be most profitable. To do so robustly, these studies should include tailor-made assessment items to measure interaction-driven learning as well as independent measures of cognitive complexity. Considering these needs, the following research questions and hypotheses guided the current study:

1. *On L2 development:* Does task complexity mediate the efficacy of recasts differently in the FTF mode compared to the CMC mode?

   Based on the predictions of the cognition hypothesis, it was hypothesized that the more complex task (+intentional reasoning) plus recasts would lead to the most L2 development. In regard to modality, given that no research has been conducted on the combined effects of complexity and mode, the null hypothesis was assumed: Modality would not affect the interaction between increased task complexity and recasts.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

2. *On tailored assessment items to measure learning:* In the cognitively complex groups (FTF+C and CMC+C), are tailored items on the assessments answered more accurately than nontailored items?

   Given the paucity of studies that have addressed this issue, a null hypothesis was adopted for this study: Tailored assessments items would not be answered more accurately than nontailored items.
3. *On measuring the construct of cognitive complexity:* Does the level of cognitive complexity of a task affect learners' reported independent measures of perceived task difficulty, regardless of mode?

   On the basis of the predictions of the cognition hypothesis, it was hypothesized that the cognitively complex task would be perceived as more difficult by learners and thus would result in (a) higher measures of perceived difficulty on the questionnaires and (b) longer time judgments in the more complex condition, irrespective of mode.

METHOD

Participants

The participants in this study were 84 undergraduate students (women, n = 46; men, n = 38) at a private university in the mid-Atlantic area of the United States. All were between the ages of 18 and 23 (M = 19.4, SD = 1.2). Seventy-seven of the participants were native speakers of English, six were native speakers of Korean, and one was a native speaker of Mandarin Chinese. Almost all engaged in some sort of CMC chat online to communicate with friends and family; many participants reported that they used Facebook chat or Gmail chat daily. All were recruited from 12 sections of a second-semester, intermediate-level Spanish course and were offered extra credit for their participation.

Operationalization of Task Complexity

The resource-directing variable operationalized for cognitive complexity in this study was +/– intentional reasoning. According to Robinson (2010), implementing this demand as a task design variable requires reflection on the intentional reasons and cognitive mental states that cause other people to do certain actions. In having to reflect on intentional reasons, learners will be encouraged to notice how matrix predicates that contain cognitive mental states involve subordination. This decision naturally led to the selection of the targeted structure in the dependent clause position, the Spanish past subjunctive, as it relates to the functional demands of +/– intentional reasoning.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Target Structure

In Spanish, the subjunctive involves two concepts: modality, or the "lexical or morphological expression of one's commitment to the truth-value of a statement" (Collentine, 2010, p. 40), and mood, the grammatical way of marking modality. Chung and Timberlake (1985; as cited in Pérez-Leroux, 2001) describe three semantic systems for the use of the subjunctive in the Spanish language: epistemic, deontic, and epistemological. The epistemological system is an attitude-based usage and was the one used in the present study. For example, verbs that denote emotion or volition (i.e., wishes, desires, needs, wants) are epistemological and take the subjunctive in the dependent clause position, whereas verbs that demonstrate the speaker's knowledge or facts take the indicative mood in a dependent clause. The subjunctive is therefore a grammatical marking on verbs in the dependent clause position as determined by the lexical content of the primary clause, which, for the epistemological system, are emotive in nature (e.g., *I'm sorry that*, *It angers me that*, etc.).

So far, research on the acquisition of the Spanish subjunctive indicates that learners acquire it in different developmental stages. The order in which adult learners initially incorporate the structure into their interlanguage appears to be mediated by the lexical class of the primary clause verb. Gudmestad (2008; as cited in Collentine, 2010) found that verbs of volition are most associated with early stages of emergence; other categories that require the subjunctive, such as uncertainty, doubt, and idiomatic expressions, are acquired later. Time reference (i.e., anticipated events) and hypotheticality may be the last to be acquired. To form the past subjunctive in Spanish, one starts with the third-person plural form of the preterit past tense. The past tense morphology is replaced with a morpheme marking mood morphology (e.g., *hablaron* "they spoke-ind" *hablaron hablara**ra**n* "they spoke-sub").[5] In this study, all participants had studied the indicative, past tense forms in Spanish. They had also learned about the concept of modality in class, had carried out tasks that required the present—but not the past—subjunctive (as mandated by verbs of volition, emotion, and anticipated events), and also had extensive practice forming subordinate clauses.

Design

A Pretest-Posttest 1-Posttest 2 design was used for this study. The 84 participants were randomly assigned to either the control group or one of the four experimental groups that differed depending on the cognitive complexity of the task (+/– intentional reasoning) and the environment

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

in which they interacted (FTF or CMC). Groups were designated according to their experimental condition: FTF+C (i.e., +intentional reasoning in the FTF mode), FTF–C (i.e., –intentional reasoning in the FTF mode), CMC+C (i.e., +intentional reasoning in the CMC mode), CMC–C (i.e., –intentional reasoning in the CMC mode), and control (i.e., no treatment, assessments only). One-way ANOVAs at the onset of the study showed that there were no statistical differences among groups with regard to age, $F(4, 83) = 1.11$, $p = .36$; sex, $F(4, 83) = 1.11$, $p = .36$; overall length of Spanish study, $F(4, 83) = 1.77$, $p = .14$; or hours per week engaging in CMC chat, $F(4, 83) = 1.91$, $p = .12$.

All participants took the pre- and posttests; the four experimental groups carried out two treatment sessions between these assessments. Second language development was measured at each testing session with an interactive task in FTF, an interactive task in CMC, and a multiple-choice receptive test delivered via the computer.

Materials

Treatment Task. The treatment task chosen for this study was an interactive dialogic story retell. Two tasks, each with a simple and a complex version, were created. The tasks satisfied the six criterial features of a task as determined by Ellis (2003) and incorporated Ellis's task design suggestions to promote L2 development as a result of interaction. They were two-way in nature, required collaborative information exchange, dealt with "human-ethical" topics (p. 96) that were familiar (as opposed to less familiar, objective topics), and had a closed outcome. The first story-retell task had to do with a family in Latin America who accused their housekeeper of stealing jewelry, only to discover that they had misplaced it. The second was about two adolescents invited to play for their city's soccer team. Both story retells were facilitated by a set of six comic cards, each preceded by a brief section of the story in English; the participant and the researcher thus each had a set of 12 cards total. In the cognitively simple condition, the intentional reasons behind certain characters' actions were already provided in the card prompts for participants. The –intentional reasoning condition required a retelling of the story events only. This is demonstrated in Figure 1, in which the intentional reason behind the main character's action is provided in both the first language (L1) blurb and in the comic strip.

In the cognitively complex condition, learners were not provided with the characters' intentional reasons in the story. They had to reflect on what caused the characters' actions by themselves, which they then had to communicate during the task. Intentional reasoning was elicited via empty thought bubbles in the comics. As shown in Figure 2,

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

L1 blurb (seen first):

> The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time, cleaned the house well, cooked good food, and cared for the children. The family really liked her and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own family was bad, but overall, she was a positive and hard working person. Every morning, Sra. Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez arrived, Sra. Martínez always smiled. She was happy that Srta. Gómez worked for them.

Comic strip card (seen second):



Translation:
Picture 1: Ms. Gómez
Ms. Gómez always …
Picture 2: Ms. Gómez (to please) us
Mrs. Martínez and her family
Picture 3: Mrs. Martínez always (to smile)
She was content that …

Figure 1. An example of the L1 story and corresponding comic card in the –intentional reasoning task.

the action for which they had to provide an intentional reason was marked with a *1*, and the empty thought bubble was marked with a *2*. Thus, where participants in the –complex condition were provided with the intentional reasons that explained the characters' actions, the +complex groups were given an empty thought bubble that prompted them to think of the intentional reasons that caused a specific action.

The use of the Spanish past subjunctive was required in both experimental conditions; the difference was that the +complex groups had to come up with the intentional reasons themselves (i.e., emotions, verbs of volition, desires, etc.) to explain actions in the story, whereas the –complex groups were already given this information. Extensive piloting with both task versions showed that this was an effective way to operationalize +/– intentional reasoning.

Assessment Tasks. To assess learners' use of the past subjunctive in communicative contexts, two productive tasks, one in FTF mode and one

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

L1 blurb (seen first):

The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time, cleaned the house well, cooked good food, and cared for the children. The family really liked her and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own family was bad, but overall, she was a positive and hard working person. Every morning, Sra. Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez arrived, Sra. Martínez always smiled.

Comic strip (seen second):



Translation:
Picture 1: Ms. Gómez
Ms. Gómez always…
Picture 2: Ms. Gómez (to please) us
Mrs. Martínez and her family
Picture 3: 1. Mrs. Martínez always (to smile)

Figure 2. An example of the L1 story and corresponding comic card in the +intentional reasoning task.

in CMC, as well as a receptive multiple-choice test, were created. Three versions of each test were created for the Pretest, Posttest 1, and Posttest 2. The productive assessment tasks were interactive in nature and were similar to the treatment tasks in that they used cards to elicit participants' retelling of a brief, real-life scenario (e.g., a couple had an argument after seeing text messages on a cell phone; a student observed her friend cheating on an exam in class). Six new stories were created for these tasks, and they were done in both the FTF and CMC modes to determine if L2 development was transferable from one mode to the other. Each story-retell task was designed to elicit 10 uses of the past subjunctive each. All six of the productive tasks were piloted on four native speakers of Spanish and on 29 second-semester, intermediate-level students of Spanish. The pilot showed that the tasks successfully elicited natural uses of the past subjunctive in each obligatory context and were comparable in lexical variation and time duration.

Three different versions of the multiple-choice test were also created, and each elicited 15 uses of the past subjunctive. The tests

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

were delivered online via Blackboard[6] so that participants could not go back and change previous answers or look at previous answers as a form of additional input. All three multiple-choice test versions were piloted on 40 additional students of second-semester, intermediate-level Spanish. To check internal-consistency reliability, Cronbach's alpha was calculated and showed that all tests were in acceptable ranges of one another (i.e., .82, .86, and .80, respectively).

For all obligatory past subjunctive contexts on the assessments, eight (in the story retells) and 10 (in the multiple choice test) of the intentional reasons were carried through from the pretest, the treatments, and the posttests. Two (out of 10) and 5 (out of 15) of the intentional reasons were reserved for tailored items in the two complex groups. Tailored items were customized to each participant in that they came from participants' own intentional reasons provided during the treatment (see the Procedure section). This was for the purpose of testing Nuevo's (2006) hypothesis that tailored items may best capture what learners contribute to the task when they must come up with reasons themselves.

All versions of the productive and receptive assessments, including the mode (i.e., FTF versus CMC) in which participants carried out the productive tasks, were counterbalanced for the assessment schedule. However, participants always took the multiple-choice test last at each testing session so as to not be prompted with input before carrying out the productive tasks.

Perceived Difficulty Questionnaire. A perceived difficulty questionnaire was created to gauge insight into learners' perception of the task's complexity. A total of 15 items were present on the questionnaire and inquired about learners' views of how hard the task was, whether or not they felt it was challenging, and their overall perception of the task's difficulty. The instrument employed a 6-point Likert scale that asked how much they agreed or did not agree with each statement (options ranged from *strongly disagree* to *strongly agree*). Cronbach's alpha was calculated to determine the internal-consistency reliability of the questionnaire; correlations were high, with .82 for the questionnaire after Treatment 1 and .88 for the questionnaire after Treatment 2.

Equipment. The equipment used for the project included (a) three Mac OS laptops, each equipped with the iChat software version 4.0 as well as iShowU HD, a screen-recording software to record all CMC interaction; (b) digital and cassette recorders to record all FTF interactions; and (c) a mobile printer to print out the tailored assessments for participants in the complex groups.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Procedure

Data were collected over the course of 1 year, with each participant attending four one-on-one sessions with the researcher. During the first session, participants were given the consent form and a background biodata questionnaire. They then carried out the pretests (i.e., two productive story-retell tasks with the researcher, one in FTF mode and one in CMC, followed by the multiple-choice receptive test). Participants then made appointments with the researcher for the remaining three sessions. The two treatment sessions took place consecutively, a maximum of 1 to 2 days apart, and occurred 1 or 2 weeks after the pretest. The control group made appointments for only the third and fourth sessions, given that they only carried out the assessments.

For the treatment sessions, participants were given instructions that explained that they would be carrying out a story-retell task with the researcher. Participants in the –complex groups (i.e., FTF–C, CMC–C) were told that they would be reading a story in sections in English and that they had to retell in Spanish each story section in the past tense as best as they could to the researcher. Participants in the +complex groups (i.e., FTF+C, CMC+C) were told the same thing but were also informed that, for some of the actions in the story, they would have to reflect on the characters' intentional reasons behind those actions. All participants were told to ask the researcher at any time during the task if they needed assistance or wanted to know how to say a word in Spanish.[7]

Interaction in the FTF mode was carried out in a room with the participant and researcher facing each other at the same table. Both the participant and the researcher started by reading the first L1 story blurb. When the participant was ready, they moved on to the first comic strip, which served as a visual prompt to help the participant retell the story in Spanish. The task was interactive and two-way in nature in that the participant and the researcher worked together to retell the story. Whenever the participant made an error with the past subjunctive, the researcher provided a full recast to correct the error. Recasts were given with a falling intonation at the end and had no emphatic added stress to them. An example from this study is provided in the example in (1) below:

(1) Participant: *Estaba molesta de que ellos buscaron en sus cosas.*
"She was annoyed that they looked-ind through her things."
Researcher: *Estaba molesta de que ellos buscaran en sus cosas.*
"She was annoyed that they looked-sub through her things."

For interaction in CMC, participants were given the same instructions as the FTF group and were told that they would be carrying out the conversation with the researcher via iChat on a Mac laptop. Unlike FTF

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

interaction, CMC interaction—both treatments and assessments—was carried out with the researcher and participant each in separate rooms (as opposed to them sitting and typing in front of each other). This was so the true nature of computerized interaction could be explored. Like those in the FTF groups, the CMC participants were told to ask the researcher if they had any questions (e.g., vocabulary items) in chat just as they would in FTF conversation; they were also told not to worry about typing accent marks. As illustrated in the example in (2), recasts were provided to learners in the CMC mode in the same way as was done in the FTF mode:

(2) Participant: *Ella dudaba que Srta. Gómez robo las perlas.*
"She doubted that Srta. Gómez robbed-ind them."
Researcher: *Ella dudaba que Srta. Gómez robara las perlas.*
"She doubted that Srta. Gómez robbed-sub them."

During the complex groups' treatment sessions, the researcher took notes on the intentional reasons that participants had to come up with during the task (e.g., "she was sad that," "he couldn't believe that," etc.).[8] Immediately after each treatment session concluded, all participants were asked to write down how many minutes they believed it took them to carry out the task they had just completed with the researcher. They then filled out the Perceived Difficulty Questionnaire.

On the second treatment day (i.e., Session 3), after completing the treatment task, the judgment of time on task, and the Perceived Difficulty Questionnaire, participants then carried out the three immediate posttests (i.e., two productive tasks in FTF and in CMC modes and then the multiple-choice receptive test). Posttest 2 was done 1 week after Session 3; afterward, all participants filled out an exit questionnaire.

The creation of tailored posttests for the +complex groups required prior planning so that the posttests could be given immediately after the second treatment with no time delay. To do this, a separate computer station with a mobile printer was set up in the next room. On this third computer, the two productive retell stories (in Word document form) that the participant was to carry out were open and ready for editing, with all 10 obligatory contexts requiring the use of the past subjunctive highlighted in yellow. Using notes taken during the treatment sessions, two of the participants' own intentional reasons were inserted into the story (replacing the two items reserved for novel items). The document was then de-highlighted, printed, and cut into card strips. This was done while participants filled out the Perceived Difficulty Questionnaire and took approximately 2–3 min; therefore, no participant in the +complex groups had to wait to begin his or her productive assessment task.

The tailored receptive multiple-choice tests were created at this same computer station while the researcher was carrying out the CMC

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

productive assessment task with the participant. The researcher replaced five of the past subjunctive items with five of the participants' own intentional reasons (once again, replacing the items reserved for novel items).

Finally, to control for outside exposure of the target form, all second-semester, intermediate-level instructors were asked to not cover the past subjunctive inside or outside of class during the experiment, and all references to the past subjunctive were removed from the syllabus and assignments.

Coding and Analyses

First, the FTF audio recordings as well as the CMC chat logs from both the treatment and production tasks were transcribed and coded for production of the past subjunctive. For the production tasks, a 0 was assigned to indicative forms, 0.5 to present subjunctive forms, and 1 to past subjunctive forms. This weighted scoring system was employed to account for developmentally sensitive L2 production (see Norris & Ortega, 2009, for arguments for coding scales representative of developmental paths in L2 acquisition).[9] Interlanguage forms (e.g., *tengaba[n], fuyera, hacera, haciera, diciera*) were assigned a full point. For the multiple-choice receptive tests, 1 point was assigned to correctly selected forms; 0 points were assigned to incorrect answers. Twenty percent of the assessment tasks and of the receptive tests were randomly selected and coded by an independent rater (another SLA researcher with 5 years of research training). Percentage agreement between the researcher's and the rater's coding was 96% or above for all of the assessment tasks (i.e., FTF or CMC, Time 1, 2, and 3) and 100% for the multiple-choice tests.

Next, descriptive statistics were calculated for (a) the average time it took to complete Treatments 1 and 2; (b) the number of recasts received per group; (c) each group's performance on the pretests and two posttests, which included performance on the tailored versus nontailored items in the complex groups; and (d) each group's average Perceived Difficulty Questionnaire score as well as time-on-task judgment as measures of cognitive complexity. To answer the research questions, a series of repeated-measures ANOVAs, Wilcoxon signed-rank tests, and factorial ANOVAs were performed (statistical models were chosen based on distribution of the data). Analyses were carried out using SPSS 19 with the alpha level set at .05. To report effect sizes, Cohen's guidelines for the interpretation of effect size magnitude were employed. For d, .20 was considered small, .50 medium, and .80 large. For r and its related indices, $r = .10$, $\eta^2 = .01$, and $R^2 = .01$ were considered small; $r = .30$, $\eta^2 = .06$, and $R^2 = .09$, medium; and $r = .50$, $\eta^2 = .14$, and $R^2 = .25$, large.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

RESULTS

Number of Recasts Provided during Treatment

An examination of Table 1 shows that the total number of recasts decreased as participants progressed from Treatment 1 to Treatment 2, albeit barely for the CMC+C group. During Treatment 1, in the FTF environment, the FTF+C group required slightly fewer recasts than the FTF–C group. These data indicate that the FTF+C group began producing the form somewhat earlier than FTF–C group. In the CMC environment, a different pattern emerged. The mean number of recasts provided to the CMC+C group stayed nearly constant, with an average of 8.88 recasts in Treatment 1, and 8.56 recasts in Treatment 2. The CMC–C group, however, received noticeably fewer recasts. The CMC–C group clearly began incorporating the form during the treatment faster than the other three groups.

L2 Development

The first research question asked whether or not task complexity worked differently in the FTF compared to the CMC mode. Before computing the descriptive statistics, three one-way ANOVAs were performed to ensure the statistical comparability of the four experimental groups and the control group at the onset of the study. The ANOVAs showed that there were no statistical differences between groups on the FTF production pretest, $F(4, 83) = 0.91$, $p = .46$; the CMC productive pretest, $F(4, 83) = 1.04$, $p = .39$; or the multiple-choice receptive pretest, $F(4, 83) = 0.59$, $p = .67$; thus, any differences between the control and the experimental groups can be attributed to the treatment. Table 2 provides the descriptive statistics (means and standard deviations) for the experimental and control groups over time; the groups' means for the FTF production

Table 1. Mean number of recasts provided during treatments per group

	Treatment 1		Treatment 2	
Group	*M*	*SD*	*M*	*SD*
FTF+C ($n = 18$)	8.61	1.94	7.83	2.60
FTF–C ($n = 18$)	9.17	0.99	8.56	1.89
CMC+C ($n = 17$)	8.88	1.54	8.56	2.55
CMC–C ($n = 17$)	6.88	2.72	4.24	3.27

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Table 2. Descriptive data for pretests and posttests

	FTF production task		CMC production task		Multiple-choice receptive test	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
FTF+C (*n* = 18)						
Pretest	0.22	0.94	0.28	0.96	2.17	1.67
Posttest 1	3.00	3.70	2.86	3.24	9.78	5.46
Posttest 2	3.17	3.72	3.61	4.31	9.33	6.40
FTF–C (*n* = 18)						
Pretest	0.00	0.00	0.06	0.24	1.39	2.25
Posttest 1	1.86	2.58	1.61	2.95	5.22	6.21
Posttest 2	1.44	2.78	1.39	2.90	4.56	5.77
CMC+C (*n* = 17)						
Pretest	0.00	0.00	0.00	0.00	2.06	3.21
Posttest 1	1.35	2.42	1.74	3.35	4.88	5.48
Posttest 2	0.59	1.30	1.12	2.63	4.82	5.49
CMC–C (*n* = 17)						
Pretest	0.00	0.00	0.06	0.24	2.06	1.25
Posttest 1	4.59	3.69	4.88	3.59	9.29	5.59
Posttest 2	4.06	3.50	4.79	3.53	10.59	5.25
Control (*n* = 14)						
Pretest	0.00	0.00	0.00	0.00	1.36	1.45
Posttest 1	0.14	0.23	0.15	0.24	2.21	2.26
Posttest 2	0.07	0.18	0.18	0.25	2.57	2.74

task, the CMC production task, and the multiple-choice receptive test are plotted in Figures 3, 4, and 5, respectively.

A visual inspection of these data shows that on all three assessments, the CMC–C and FTF+C groups outperformed the other three groups on each assessment at Posttest 1 and Posttest 2. On both of the production tasks, the CMC–C group performed the best, followed closely by the FTF+C group. The FTF+C group was the only group to continue improving at Posttest 2 for the productive tasks. The CMC+C group performed the worst. On the multiple-choice receptive test, the FTF+C group performed the highest at Posttest 1 and was followed closely by the CMC–C group. Three separate 3 × 2 × 2 repeated-measures ANOVAs were then performed to see if the differences in the effects of task complexity and mode on participants' scores were statistically significant; the results are reported for each assessment.

FTF Production Task. Results from the ANOVA revealed a statistical interaction between the two independent variables of complexity and

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

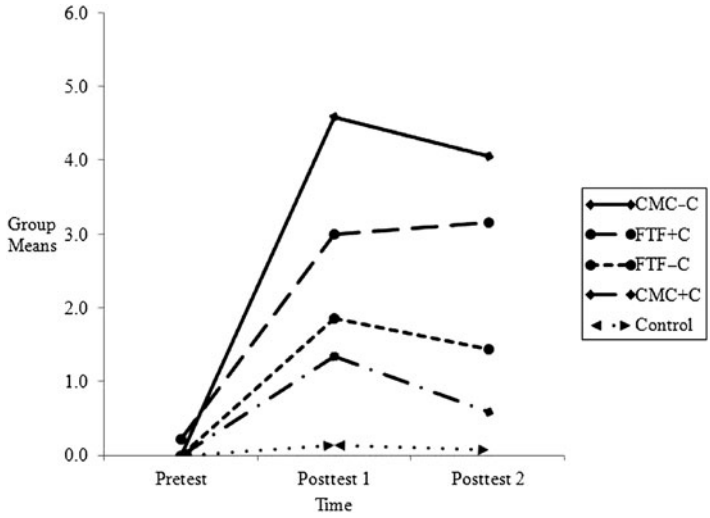


Figure 3. Pre-to-post development on the FTF production task.

modality, $F(1, 66) = 13.20$, $p = .001$, partial $\eta^2 = .17$, power = .95, as well as a statistical triple interaction between time, complexity, and modality, $F(1.99, 131.18) = 7.28$, $p = .001$, partial $\eta^2 = .10$, power = .93. Effect sizes for the interaction effects were high, and the power analysis showed that both analyses had sufficient power (more than 90%) to find statistical differences in the interaction. This indicated that on the FTF production assessment task, learners performed differently with tasks

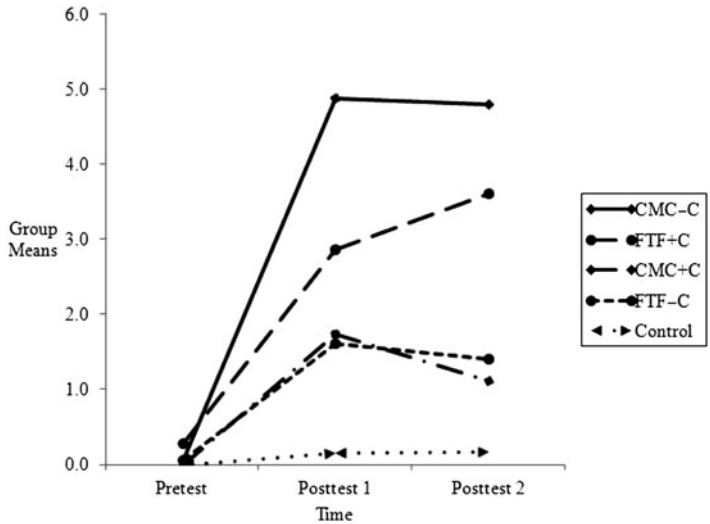


Figure 4. Pre-to-post development on the CMC production task.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

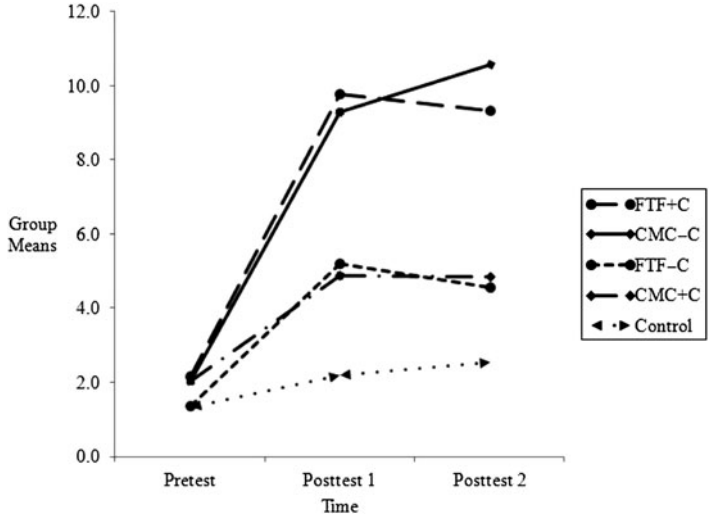


Figure 5. Pre-to-post development on the multiple-choice receptive test.

of increased cognitive complexity depending on the interaction environment in which they carried out the task. Performing the cognitively complex task in the FTF mode resulted in statistically more L2 development on the FTF production task. Contrarily, performing the cognitively simple task in the CMC mode resulted in statistically more L2 development.

CMC Production Task. The ANOVA revealed a statistical interaction between complexity and modality, $F(1, 66) = 12.13$, $p = .001$, partial $\eta^2 = .16$, power = .93, and also a statistical triple interaction for time, complexity, and modality, $F(1.84, 121.5) = 7.97$, $p = .001$, partial $\eta^2 = .11$, power = .94. Both analyses had large effect sizes. For learning as measured by the CMC production assessment task, cognitive complexity was once again found to mediate L2 development, and modality affected this outcome. Performing the cognitively complex task in the FTF mode resulted in statistically more L2 development on the CMC production task than the simple task; L2 development continued to improve in this condition at Posttest 2. For learners who interacted in the CMC mode, performing the cognitively simple task led to the most L2 development, as measured by the CMC production task.

Multiple-Choice Reception Test. The ANOVA showed a statistical interaction between complexity and modality, $F(1, 66) = 12.96$, $p = .001$, partial $\eta^2 = .16$, power = .94, as well as a highly statistical triple interaction among time, complexity, and modality, $F(1.69, 111.6) = 9.93$, $p < .001$,

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

partial η^2 = .13, power = .97. The effect sizes for both interactions were high. Modality once again made a difference on the effects of cognitive complexity and subsequent L2 development. Carrying out the cognitively complex task in FTF mode led to the most L2 development as measured by the multiple-choice test (with mean scores almost double that of the group that performed the cognitively simple task in the same mode). Carrying out the cognitively simple task in CMC led to the most L2 development, and this condition resulted in even higher scores at Posttest 2.

These findings indicate that task complexity worked differently depending on the environment in which interaction took place. When learners interacted and received recasts, carrying out the more cognitively complex task assisted them in accurately marking past tense verbs with subjunctive morphology but only in the FTF mode. In the CMC mode, carrying out the cognitively simple task resulted in the highest mean scores for accurately marking past tense verbs with subjunctive morphology on both of the productive tasks as well as the multiple-choice receptive test.

Measuring Learning with Tailored versus Nontailored Items

The second research question asked whether tailored items (i.e., those intentional reasons that participants in the complex groups came up with themselves) on the assessments would be answered more accurately than nontailored items. A tailored score and a nontailored score were computed for each assessment at Posttest 1 and Posttest 2 (i.e., FTF productive, CMC productive, and multiple-choice receptive) for participants in the FTF+C and CMC+C groups. A series of Wilcoxon signed-rank tests were conducted to examine if differences between how participants answered tailored versus nontailored items were statistical. Mean ranks, *z* scores, *p* values, and effect sizes are reported in Table 3.

The Wilcoxon signed-rank tests revealed statistical differences between tailored and nontailored items on the productive (FTF and CMC) Posttest 1, with tailored items resulting in learners' production of past subjunctive statistically more often than did those items that were not tailored. The statistical difference on both assessments had a large effect size (r = .5). On Posttest 2, these differences remained statistical: Tailored items resulted in the production of the past subjunctive statistically better than nontailored items, and the effect size for these differences was medium (r = .4). For both of the productive assessment tasks (FTF and CMC), whether or not the participant came up with the intentional reason—versus provision of the reasoning to the

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Table 3. Descriptive statistics for tailored versus nontailored assessment items

Assessment	Item type	*N*	*M* rank	*z*	*p* value	Effect size
FTF Posttest 1	Tailored	35	9.85			
	Nontailored	35	2.67	–3.11	.002	*r* = .5
FTF Posttest 2	Tailored	35	7.50			
	Nontailored	35	2.00	–2.40	.016	*r* = .4
CMC Posttest 1	Tailored	35	11.38			
	Nontailored	35	3.30	–2.85	.004	*r* = .5
CMC Posttest 2	Tailored	35	9.78			
	Nontailored	35	3.40	–2.24	.025	*r* = .4
M-C Posttest 1	Tailored	35	11.31			
	Nontailored	35	11.78	–0.67	.502	*r* = .1
M-C Posttest 2	Tailored	35	8.66			
	Nontailored	35	14.50	–2.96	.003	*r* = .5

Note. M-C = Multiple-choice

participants—made a difference: Participants produced the past subjunctive in the dependent clause significantly better following their own intentional reasons than for ones they did not come up with themselves. On the multiple-choice receptive test, results were slightly different. There was no difference in how participants answered tailored versus nontailored items on the recognition test. In fact, mean rank was nearly identical on Posttest 1: Tailored items = 11.31, nontailored items = 11.78. At the multiple-choice receptive Posttest 2, participants answered nontailored items statistically more accurately than tailored items (mean rank was 8.66 and 14.50, respectively), with a large effect size (r = .5). On the multiple-choice test, there was no difference in how participants answered the assessment items, and, in fact, nontailored items were answered statistically better at Posttest 2.

Independent Measures of Cognitive Complexity

The third research question asked whether or not the cognitive complexity level of a task affected participants' reported independent measures of perceived task difficulty and if this effect held for both modes. The dependent variable measures came from participants' Perceived Difficulty Questionnaire (Treatment 1 and 2) and their retrospective time-on-task judgments (Treatment 1 and 2). For the questionnaires, an overall perceived difficulty score was computed for each participant for each treatment session. For the time-on-task judgments, a time difference score was calculated for each participant, in which participants'

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

real treatment time was subtracted from their guessed time for both treatment sessions. Two 2 (Complexity) × 2 (Modality) factorial ANOVAs were performed on the dependent variables, one for each treatment session. Results are reported for each variable.

Perceived Difficulty Questionnaire. The descriptive statistics for learners' reported perceived difficulty during the experiment are reported in Table 4. A visual analysis of group mean scores show that the FTF+C group had the highest perceived difficulty scores, followed by FTF–C, CMC+C, and then CMC–C at both treatment sessions. It appeared that modality, as opposed to task complexity, affected this dependent variable more.

The results from the ANOVA confirmed this observation and did not reveal a statistical effect for the main effect of complexity, $F(1, 66) = 1.97$, $p = .17$, partial $\eta^2 = .03$. A statistical effect was found, however, for the main effect of modality, $F(1, 66) = 10.97$, $p = .002$, partial $\eta^2 = .14$, with a large effect size. The interaction of the two main effects was not statistical. During the first treatment, participants rated the task as more difficult if they interacted in the FTF environment, regardless of whether or not the task was cognitively complex. Adjusted R^2 showed that the ANOVA accounted for 12.9% of the variance in the data at Treatment 1 (a medium effect) with power at 85.6%. These same results were found at Treatment 2. There was no statistical effect for complexity, $F(1, 66) = 3.39$, $p = .07$, partial $\eta^2 = .05$, but a statistical effect for modality was found and with a medium effect size, $F(1, 66) = 8.58$, $p = .005$, partial $\eta^2 = .12$. Once again the interaction between complexity and modality was not statistical. The effect size for the main effect of modality at Treatment 2 remained at a medium level, accounting for adjusted $R^2 = 11.6\%$ of the variance in the data with a power of 81.7%. Therefore, the cognitive complexity level of the task did not affect learners' perceived difficulty of the task as measured by a questionnaire. Instead, the environment in which participants interacted statistically affected their perceived

Table 4. Perceived task difficulty according to group as measured by the questionnaire

	Treatment 1		Treatment 2	
Group	*M*	*SD*	*M*	*SD*
FTF+C ($n = 18$)	49.38	9.38	48.25	9.02
FTF–C ($n = 18$)	45.27	7.36	43.28	11.1
CMC+C ($n = 17$)	41.24	8.60	40.71	10.3
CMC–C ($n = 17$)	39.53	9.64	37.09	8.60

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

difficulty, and it did not matter if the task was complex or simple. The FTF groups always rated the task as more difficult than the CMC groups, regardless of whether the task was simple or complex. Although the complex tasks were rated as slightly more difficult in both modes, the difference compared to simple tasks was not statistical.

Retrospective Time-on-Task Judgments. The descriptive statistics for participants' mean judgments of how much time they spent on the task compared to their actual time on task are provided in Table 5. A visual inspection of the data indicated that learners who performed the complex task—regardless of whether they interacted in FTF or CMC mode—judged their time on task as greater than their real time. On average, those who carried out the simple task judged that task as taking less time than their real time. Two 2 × 2 factorial ANOVAs examining the effects of complexity and modality on how differently participants judged their time on task from their real time found a statistical effect for complexity, $F(1, 66) = 33.99$, $p < .001$, partial $\eta^2 = .34$, at Treatment 1; the effect size was very large. The main effect of modality was not statistical, $F(1, 66) = .004$, $p = .95$, partial $\eta^2 = .00$; nor was the interaction effect statistical, $F(1, 66) = 3.24$, $p = .08$, partial $\eta^2 = .05$. Carrying out the cognitively complex task (i.e., having to intentionally reason during the interaction) resulted in participants judging the task as taking significantly more time than it really did; those who performed the cognitively simple task judged it as taking significantly less than the real time (at Treatment 1, the cognitively complex groups guessed an average of +2.89 min in the FTF mode and +5.82 min in the CMC mode, whereas the simple groups guessed an average of –4.00 min in FTF mode and –7.15 min in CMC). The effect size showed that the ANOVA accounted for $R^2 = 33.2\%$ of the variance in time judgments—a high effect—and also with high power (1.0). Increased cognitive complexity led to statistically higher time-on-task judgments, and this independent measure held across mode. At Treatment 2, both of the main effects of complexity, $F(1, 66) = 7.72$, $p = .007$, partial $\eta^2 = .11$, and modality, $F(1, 66) = 6.03$, $p = .02$, partial $\eta^2 = .08$, were statistical, though their interaction was not, $F(1, 66) = 1.08$, $p = .30$, partial $\eta^2 = .02$. Adjusted R^2 showed that the ANOVA model accounted for 14.6% of the variance in time judgments, with a power of 89.6%. This finding indicated that at Treatment 2, those participants who carried out the complex task always judged it as taking more time than the real lapsed time, and, additionally, those in the FTF mode judged the task as taking more time than the CMC mode.

In sum, the two independent measures of cognitive complexity—the Perceived Difficulty Questionnaire and retrospective judgments of time on task—revealed different results about learners' perception of cognitive complexity. The questionnaire measure showed no effect for task complexity. Rather, the environment in which participants interacted

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Table 5. Time judgments (in minutes) of guess time compared to real time

	Treatment 1						Treatment 2					
	Real time		Guess time		Difference		Real time		Guess time		Difference	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
FTF+C	25.17	4.90	28.06	9.30	+2.89	8.53	21.83	5.73	24.50	10.45	+2.67	9.25
FTF–C	21.17	6.96	17.17	5.77	–4.00	5.38	16.50	5.47	16.28	6.87	–0.22	4.05
CMC+C	58.00	13.62	63.83	15.76	+5.82	6.55	48.65	14.04	49.00	12.18	+0.35	7.04
CMC–C	43.47	8.84	36.32	11.04	–7.15	7.41	33.00	7.66	27.06	7.92	–5.94	5.92

statistically affected this measure, with those in the FTF mode rating the task statistically more difficult than those in the CMC mode. A statistical effect for complexity was found on the time judgment measure. Regardless of the environment in which they interacted, participants who carried out the complex task judged it as having taken significantly more time than the real time they needed to perform the task; those who carried out the simple task judged it as taking significantly less time than their real time. By Treatment 2, an effect for modality on time judgments was also found, but it was a smaller effect than cognitive complexity.

DISCUSSION

Two key claims of Robinson's (2007, 2011) cognition hypothesis were tested in this study. The first is that increasing the cognitive complexity of a task will result in learners allocating greater attention to and increased memory of forms provided via reactive focus-on-form techniques, such as recasts, during conversational interaction (Robinson, 2011). The second is that differences in the cognitive complexity of a task will be matched by learner perceptions of task difficulty. As the cognition hypothesis predicts, this study found that increases in task complexity resulted in more learning, but in the FTF mode only. Engaging in intentional reasoning while interacting and receiving recasts led to more L2 development in the FTF mode; this was not the case in CMC. Rather, performing the cognitively simple task led to the most L2 development in CMC. The modality in which learners interacted therefore mediated the effects of cognitive complexity. Looking at independent measures of cognitive complexity, the present study also found that learners' perceptions of task difficulty matched the researcher operationalization of increased cognitive complexity on only one of the measures: retrospective judgments of time on task. There was no effect for increased cognitive complexity as measured by the Perceived Difficulty Questionnaire.

These findings indicate that task design can affect the efficacy of recasts. It is possible that corrective feedback may work best congruently with more cognitively complex tasks in the FTF mode. In the CMC mode, it may be the case that feedback in the form of recasts works best with simple tasks that do not pose a cognitive overload. Having to engage in intentional reasoning appeared to push those learners in the FTF+C group to process the recasts they received at a deeper level, whereas those in the FTF–C were not as compelled to do so. In contrast, recasts alongside cognitively more complex tasks in the CMC mode may have been too much for learners to process.[10] This outcome suggests that, for future studies, task complexity may need to be considered as a variable with the capacity to mediate the efficacy of recasts and, critically,

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

differentially so according to the environment (i.e., traditional versus online classes) in which a L2 is being taught.

One question that merits further investigation is why cognitive complexity was experienced so differently in FTF versus CMC modes. To explore this result in greater detail, a follow-up analysis was carried out looking at the qualitative nature of interaction and feedback provision for each experimental condition. This was done by reviewing all interaction from the FTF audio recordings and the CMC screen-recording videos, which revealed interesting features about the provision and reception of recasts alongside cognitively simple versus complex tasks.

Performing the cognitively simple task in CMC necessitated fewer turns to communicate the main points of the story than performing the cognitively complex task, and the amount of discourse per turn was also shorter in length. When participants in the CMC–C group received corrective recasts, the timing of the feedback was often immediately after their erroneous production. The visual comparison of the problematic utterance with the recast showed two almost identical utterances; the only difference was the corrected form in the recasted utterance. The screen-recorded videos also showed that these participants were typically waiting for a response from the researcher before typing out their next message and were therefore not engaged in any other activity when the feedback arrived. Equally revealing is the fact that some participants in CMC–C attempted to write out the past subjunctive, then moved their mouse upward or scrolled up to examine formerly provided feedback, and then sent their message. This is demonstrated by the participant in (3) below:

(3) ~~tuvo~~ ~~tuvieron~~ tuviera [scrolled up to confirm form] tuvieran [sent message]

This participant wrote the past tense indicative and singular form of the verb *tener* "to have": *tuvo*. He then erased it (represented here as crossed-out text) and typed out the plural verb form, still in the indicative: *tuvieron* "they had." He then erased the indicative morphological ending and replaced it with subjunctive morphology. Before sending the message, the participant scrolled up quickly to view a past subjunctive form (from a previous recast) for comparison. After seeing that it was correct, he scrolled back down, added the plural morpheme *-n*, and then sent the message to the researcher. This example indicates that learners in the CMC–C group had more time as well as more attentional resources available during the task to test their hypotheses about a form and confirm the accuracy of their production.

Performing the cognitively complex task in CMC involved turns that were notably longer and confusing, which led many participants in the CMC+C group to apologize during the sessions. The issue of split negotiation routines as described by other researchers (e.g., Lai et al., 2008)

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

was evident in this condition: At times, some participants provided the intentional reason (the primary clause) and its subordinate clause (that required the subjunctive) in nonadjacent turns in the chat. Sometimes, it took up to 10 turns just to get participants to reflect on the intentional reasoning of the characters. The erroneous use of the indicative past tense and subsequent repair that made up recasts in CMC+C were often not contingent in time, rendering the juxtaposition of the recasts after erroneous production problematic and even delayed. Additionally, an examination of the screen-recorded videos revealed that participants in CMC+C were almost always typing their next message when a recast arrived from the researcher. Turn taking in CMC+C was problematic precisely because of this, because both the participant and the researcher were sending or typing another message simultaneously. This may explain why carrying out the cognitively complex task in CMC often led to cognitive overload and frustration and, in some cases, caused participants in this group to miss the feedback entirely.

In the FTF–C group, turn taking during the conversations was brief and straightforward. Carrying out the cognitively simple task in FTF mode required relatively little scaffolding to complete the task. The participants were provided with the characters' intentional reasons behind their actions in the story blurb and comics and had to relay the contents of the story. The researcher's corrective recast followed immediately after each erroneous production of the past subjunctive, and acknowledgement of the feedback (e.g., *sí* "yes") as well as the production of immediate (often exact) uptake was omnipresent in this condition. This is demonstrated in the example in (4) by the way in which the following participant reacted to a recast:

(4) Participant:
Sí, y ella quería que Srta. Gómez regresó a la casa.
"Yes and she wanted [that] Ms. Gómez return-ind to the house."
Researcher:
Ella quería que Srta. Gómez regresara a la casa.
"She wanted [that] Ms. Gómez return-sub to the house."
Participant:
. . . regresara a la casa, uh huh, porque, quería decir "lo siento."
". . . returned-sub to the house, uh huh, because, she wanted to say 'I'm sorry.'"

The participant repeated the corrected element of recast perfectly and then quickly went on to make another point.

In FTF+C, turns in the discourse were much longer and involved a greater amount of scaffolding and negotiation of meaning, especially with regard to requests for vocabulary of feelings or other cognitive terms on behalf of the participants. As with erroneous production in the FTF–C group, recasts were provided to the FTF+C learners right after their errors,

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

which set up an immediate juxtaposition of the two utterances (erroneous vs. correct) so that a cognitive comparison of the forms could be carried out. However, having to reflect on the intentional reasons that caused the characters' actions appeared to push learners in FTF environments to be more receptive of corrective feedback. Whereas participants in FTF–C quickly imitated the feedback and moved on, many participants in the FTF+C group seemed to assimilate the feedback more by repeating the linguistic form slowly and by highlighting the new subjunctive morphology that replaced their own indicative mood marking. This notion is demonstrated by the way in which one participant in FTF+C repeated and produced partial uptake during the task: "*tuviera . . . tuverie . . . RA*" (here, the participant repeated the form and then repeated the form again, this time isolating the subjunctive marking on the verb with a raised intonation). Some participants in FTF+C even made metalinguistic comments to themselves during the task ("*El subjuntivo . . .*" ["The subjunctive . . ."]). The fact that the intentional reasons were the participants' own reflections may have been a promoter of development: All interaction and negotiation work during the task was based on what the learner came up with, making the task more learner relevant. These online data suggest that feedback alongside more cognitively complex tasks in the FTF mode pushed learners to hold on to recasts longer and to process them more. For the most part, the learners in the FTF–C group did not demonstrate this while performing the task.

It appears, then, that the benefits afforded by task complexity in enhancing the impact of recasts are differentially mediated by the environment in which learners interact. In this study, cognitive complexity induced a greater incorporation of feedback during interaction in the FTF mode. Increasing the cognitive demands of tasks in the CMC mode seemed to have eradicated those benefits of CMC that are hypothesized to promote attention to form. This is a key issue to investigate, especially as more language courses incorporate technology and CMC environments. If, as the cognition hypothesis (e.g., Robinson, 2007, 2010, 2011; Robinson & Gilabert, 2007) predicts, performing tasks of increased cognitive complexity better equips learners to carry out authentic tasks in the real world, then the question is raised with regard to the overall utility of CMC-based interaction and feedback alongside tasks that, in FTF environments, are theorized to maximize L2 acquisition. Practitioners must consider the implications for technology in foreign language education and be more critical of where it is most effective and where it is not. Although it could be argued that cognitively simple tasks in CMC are a means to lessen the burden on those attentional resources needed to cognitively compare recasts (Doughty, 2001), CMC's capacity to prepare learners for real-life language tasks has to be questioned if more evidence is found showing that it is not a suitable environment for performing (and learning from) sociocollaborative tasks that are cognitively complex.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Another important issue to consider is that of transferability of L2 development from one mode to the other. In this study, transferability of learning was attested in three out of the four experimental groups. A quick comparison of development shows that participants who performed the experiment in the FTF mode (both the FTF+C and FTF–C groups) as well as the CMC–C group achieved some gains on the FTF *and* CMC postassessments, which indicates a transfer of development from FTF to CMC modes. However, the CMC+C group achieved a mean score of only 0.58 on the delayed FTF productive assessment, so it is not clear if learning in this group transferred to the FTF mode. This fact reiterates the finding that cognitive complexity may not work in CMC.

This study also found differences in the way in which learners produced the targeted form after tailored versus nontailored items on the postassessments. Learners in both of the complex groups were statistically better at producing (or, for the multiple-choice test, at selecting) the Spanish past subjunctive in subordinate clauses following their own intentional reasons. This was an important finding because it showed that, as Nuevo (2006) predicted, tailored assessments may be better for measuring learning that arises from task-based interaction with complex tasks, given that they acknowledge what the learner contributed to the task.

Lastly, the present study found that questionnaires did not reveal a match between the researcher's rendition of cognitive complexity and participants' perception of the construct. Rather, participants' time judgments were most revealing with regard to cognitive complexity. Those participants who carried out the cognitively complex tasks judged the task as taking significantly more time than the time it actually took. Learners who did not have to engage in intentional reasoning judged the tasks as taking significantly less time than had passed. On the basis of results from the present study, subjective time estimations may be one way to confirm that manipulations of cognitive complexity in task design have been achieved and thereby to validate cognitive complexity as an independent variable. To reaffirm Norris and Ortega's (2009) arguments, this finding certainly underscores the need for appropriate measures of cognitive complexity that are independent of linguistic output.

LIMITATIONS AND FUTURE RESEARCH

The present study is not without limitations. First, the linguistic item investigated may limit the generalization of this study's findings. The Spanish past subjunctive can be a difficult form to acquire from conversational interaction, given that it requires (a) the ability to select which mood to use in a subordinate clause, (b) knowledge of the subjunctive

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

morphological inflections, and (c) the ability to process information (i.e., modality and then mood) between a main and a subordinate clause (Collentine, 2010). Future studies would benefit by exploring the acquisition of other linguistic items in task-based interaction. This will also contribute to the body of research on the effectiveness of recasts and how the saliency of the form(s) they target makes a difference.

Second, the study only examined aggregated group data. As pointed out by one anonymous reviewer, looking into individual data could have provided a different lens for the results, as recent research on acquisition of L2 as a complex adaptive system has argued (e.g., N. C. Ellis & Larsen-Freeman, 2009; Larsen-Freeman, 2006). Therefore, a follow-up study should also examine within-group data. Additionally, the FTF+C group continued to improve on both of the FTF and CMC production tests after the treatment sessions ceased, whereas all other groups showed a loss of learning at Posttest 2. On the multiple-choice receptive test, the CMC–C group was the only one that showed continued posttreatment development at Posttest 2. These results are in line with other studies that have suggested that the effects of feedback may be delayed (e.g., McDonough & Mackey, 2006), and a true delayed posttest could have provided further insight into this trend. Future designs would profit from looking at group as well as individual variation and from longitudinal designs with more delayed assessments.

Third, only two modes of interaction were explored in this study. It would be interesting to see what roles other modes, such as video Skype (which many of the participants in this study reported using often), play in task-based interaction. Other follow-up studies could also examine different types of focus-on-form techniques during interaction and could look at how task design and interaction environment mediate their efficacy. Diverse forms of assessments to measure interaction-driven learning must also be utilized. For example, more research is needed that employs different methods to assess L2 development as it arises from task-based interaction with complex tasks, looking at tailored and standardized assessments. This will require observing learners as they complete the task, so that researchers can record learner-derived reasons, intentions, emotions, perspectives, and even linguistic forms. Doing so will improve the validity of assessments that can successfully measure interaction-driven learning as it arises from complex tasks. Until now, the extent to which CMC learning translates into FTF performance abilities, and, to a lesser extent, the reverse, has not been well established in the field. To test cross-modal transfer effects, it is also critical that future studies employ assessment tasks in both modes and examine the effects of task design features that moderate these outcomes.

Finally, and especially for studies that operationalize cognitive complexity, more empirical ways of measuring the construct that are

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

independent of linguistic complexity (from oral or written production) are needed. Given that this study was the first to use retrospective time judgments to measure cognitive load and task complexity, more research that uses time judgments as well as other independent measures of the construct are necessary for validation purposes. Doing so will improve empirical efforts to examine how task complexity differentially affords opportunities for L2 acquisition and in different interactional settings as well as the role it plays for SLA theory.

Received 12 September 2011
Accepted 26 June 2012
Final Version Received 2 August 2012

NOTES

1. Language-related episodes are moments during interaction when learners "talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326).

2. There is an extensive body of literature in psychology that shows that retrospective judgments of time on task serve as a function of the amount of information processed, indicating that they are one and the same dimension. Cognitive load is defined as the "information-processing (attentional or working memory) demands" of a task (Block et al., 2010, p. 330), and Block et al.'s (2010) meta-analysis showed that "duration judgments" serve as a means to properly measure this construct.

3. One of the reviewers asked if time estimation could be affected by the amount of recasts provided in the interaction, which is a task-external factor. However, the fact that the FTF–C group received more recasts than the FTF+C group but judged less time having passed than did the FTF+C group indicates that the number of recasts provided did not affect this dimension.

4. Web 2.0 refers to the second generation of Internet usage, which is focused on interaction and collaboration (e.g., social networks) as well as user-created content.

5. Here and throughout, ind refers to indicative marking; sub refers to subjunctive marking.

6. Blackboard is an education and course management software; participants were observed going back to previous items on paper test versions during the pilot study.

7. Participants were not given planning time before carrying out the tasks.

8. Note-taking was done while the participant went on to read the next story card so as not to interfere with the two-way interaction during the story retell component.

9. Marking a verb with the subjunctive morphology—even if in the present tense—while retelling a story could be indicative of the fact that the learner recognized that the verb needed mood marking but did not know how to do this in the past tense. U-shaped behavior observed during the treatment supported this theory.

10. An anonymous reviewer commented that the fact that the +complex group did so poorly in CMC could be considered a trade-off of attentional capacities, as Skehan (2009) has described. However, although the trade-off hypothesis and cognition hypothesis differ with regard to their predictions on the complexity, accuracy, and fluency (CAF) of learners' L2 production, the cognition hypothesis makes very specific and additional predictions about learners' noticing of form after focus-on-form techniques, uptake, and learners' perception of increased complexity. It is these predictions that were operationalized and studied in this article. Although discussion of the two theories' approaches to CAF exceeds the scope of the present study, the data will be analyzed within the theoretical framework of the trade-off hypothesis in a follow-up study.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

REFERENCES

Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study. *Modern Language Journal*, *88*, 31–55.

Beauvois, M. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals*, *25*, 455–464.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, *4*, 120–136.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*, 330–343.

Chapelle, C. (1998). Analysis of interaction sequences in computer-assisted language learning. *TESOL Quarterly*, *32*, 753–757.

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. New York: Cambridge University Press.

Collentine, J. (2010). The acquisition and teaching of the Spanish subjunctive: An update on current findings. *Hispania*, *93*, 39–51.

de la Fuente, M. J. (2003). Is SLA interactionist theory relevant to CALL? A study on the effects of computer-mediated interaction in L2 vocabulary acquisition. *Computer Assisted Language Learning*, *16*, 47–81.

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

Doughty, C. J., & Long, M. H. (2003). *The handbook of second language acquisition*. Oxford: Blackwell.

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

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, *59*, 90–125.

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

Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *CALICO Journal*, *19*, 279–294.

Fink, A., & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence, and time estimation as an index of cognitive load. *Personality and Individual Differences*, *30*, 1009–1021.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, *47*, 367–395.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34*, 445–474.

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

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System*, *37*, 254–268.

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition*, *34*, 627–658.

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 287–306). Amsterdam: Benjamins.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics*, *45*, 261–284.

Lai, C., Fei, F., & Roots, R. (2008). The contingency of recasts and noticing. *CALICO Journal*, *26*, 70–90.

Lai, C., & Zhao, Y. (2006). Noticing in text-based online chat. *Language Learning & Technology*, *10*, 102–120.

Lamy, M.-N., & Hampel, R. (2007). *Online communication in language teaching and learning*. Basingstoke, UK: Palgrave Macmillan.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*, 590–619.
Lee, L. (2001). Online interaction: Negotiation of meaning and strategies used among learners of Spanish. *ReCALL*, *13*, 232–244.
Long, M. H. (2000). Second language acquisition theories. In M. Byram (Ed.), *Encyclopedia of language teaching* (pp. 527–534). London: Routledge.
Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Erlbaum.
Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). New York: Cambridge University Press.
Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*, 37–66.
Mackey, A. (2007). *Conversational interaction in second language acquisition*. Oxford: Oxford University Press.
McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning*, *56*, 693–720.
Michel, M. (2011). *Cognitive and interactive aspects of task-based performance in Dutch as a second language* (Unpublished doctoral dissertation). University of Amsterdam, Netherlands.
Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719–758.
Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578.
Nuevo, A. (2006). *Task complexity and interaction: L2 learning opportunities and interaction* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.
Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71.
Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence in the virtual foreign language classroom. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 59–86). New York: Cambridge University Press.
Pérez-Leroux, A. (2001). Subjunctive mood in Spanish child relatives: At the interface of linguistic and cognitive development. In K. Nelson, A. Aksu-Koç, & C. Johnson (Eds.), *Children's language* (pp. 69–93). Mahwah, NJ: Erlbaum.
Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, *25*, 99–126.
Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437–470.
Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *Modern Language Journal*, *95*, 168–181.
Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). New York: Cambridge University Press.
Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.
Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, *45*, 191–213.
Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Inside the learner mind: Cognitive processing in second language acquisition* (pp. 239–264). Amsterdam: Benjamins.
Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 3–38). Amsterdam: Benjamins.
Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching*, *45*, 161–176.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press

Sachs, R., & Suh, B.-R. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 197–227). Oxford: Oxford University Press.

Salaberry, R. (2000). L2 morphosyntactic development in text-based computer-mediated communication. *Computer Assisted Language Learning*, *13*, 5–27.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke, UK: Palgrave Macmillan.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, *13*, 96–120.

Sauro, S., & Smith, B. (2010). Investigating L2 performance in text chat. *Applied Linguistics*, *31*, 554–577.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510–532.

Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly*, *39*, 33–58.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, *82*, 320–337.

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171–195). Oxford: Oxford University Press.

https://doi.org/10.1017/S0272263113000429 Published online by Cambridge University Press