Task Completion and its Effects on Linguistic Complexity/Accuracy in a 3D World

El cumplimiento de las tareas y sus efectos sobre la complejidad y precisión lingüística en un mundo 3D

KARINA COLLENTINE
Department of Global Languages
Northern Arizona University

Abstract

Task-based language teaching (TBLT) researchers argue that tasks' outcomes and efficacy should not only consider learners' performance (e.g., linguistic complexity, accuracy, and fluency) but also task completion (i.e., if learners attain a task's 'communicative objectives'), since the two processes operate symbiotically (Kuiken *et al.* 2010; Pallotti, 2009). This study provides insights into the relationship between the production of linguistic complexity and accuracy and the attainment of communicative objectives in a task entailing instant messaging. Third-year L2 learners of Spanish (N = 66) participated in two tasks, each containing a 3D world exploration segment and a subsequent synchronous computer-mediated communication (SCMC) segment. Quantitative and qualitative analyses assessed the relationship between learners' production of four variables representing linguistic complexity and accuracy in the SCMC segment and whether they completed the tasks. The analysis indicates that the extent to which learners attain a task's communicative goals interacts with their production: learners who achieved the tasks' communicative objectives produced discourse containing numerous clauses per c-unit but with numerous errors. Conclusions address the importance of considering complexity/accuracy and task completion as interacting constructs, suggestions for future CALL research, and pedagogical implications.

Keywords: task-based language teaching, computer-assisted language learning, linguistic complexity, task completion, 3D

Resumen

Los expertos en el aprendizaje a base de tareas razonan que la eficacia de una tarea debe considerar no solo la actuación del aprendiz (ej. la complejidad lingüística, la precisión gramatical y la fluidez) sino también la culminación de la tarea (i.e., si el aprendiz logra el objetivo comunicativo), ya que trabajan en simbiosis (Kuiken *et al.,* 2010; Pallotti,

2009). El presente proyecto investiga la relación entre la producción de la complejidad lingüística y la precisión gramatical, y el logro del objetivo comunicativo de dos tareas; cada una con dos partes: la exploración de un mundo virtual seguida por un segmento de comunicación mediada por computadora. Participaron 66 aprendices del español como L2 que cursaban el tercer año de estudios universitarios. Los análisis evaluaron la relación entre la producción de cuatro variables que representan la complejidad lingüística y la precisión gramatical en el segmento de CMC y la culminación de las tareas. El análisis indica que existe una interacción entre la culminación y la producción: aquellos que lograron el objetivo produjeron numerosas cláusulas por unidad comunicativa pero a la vez muchos errores. Se describe la importancia de considerar la interacción de estos dos constructos, sugerencias para la investigación futura e implicaciones pedagógicas.

Palabras clave: el aprendizaje a base de tareas, el aprendizaje mediado por computadora, la complejidad lingüística, la culminación de una tarea, 3D

Introduction

One of the goals of second language acquisition (SLA) research is to delineate the most effective types of activities that will give language learners opportunities to produce discourse containing linguistic complexity, as Skehan (1996) posits that processing complex language is more important to language development than accuracy or fluency. Activities that follow task-based instructional principles are thought to be particularly useful in fostering linguistic complexity since they lead to the development of complexity, along with accuracy and fluency (CAF) (Robinson, 2001). Such tasks have the following characteristics: "meaning is primary and [they] have a relationship to the real world; task completion has some priority; and the assessment of task performance is in terms of task outcome" (Skehan, 1996, p. 38).

Such definitions of tasks tend to dichotomize 'task completion' –whether learners attain a task's 'communicative objectives'– and 'task performance' –the nature of the language learners produce– as if they were constructs to be assessed independently, although there is reason to suspect that the constructs interact symbiotically, such that an assessment of linguistic complexity resulting from a task (or lack thereof) should consider whether the learner actually completed the task's nonlinguistic, communicative objectives (Pallotti, 2009). Indeed, Pallotti (2009) posits that learners focused on meeting a task's communicative objectives could generate CAF in unexpected ways. He notes that an utterance may be complex (e.g., *Es posible que alguien lo haya logrado asesinar* 'It is possible that someone has managed to kill him') but only marginally helpful towards meeting a task's objective (i.e., determining who committed a crime). Another utterance may have little linguistic complexity and low accuracy (e.g., *Yo *creer que Juan mata *lo* 'I *to believe that Juan *him kill') and yet constitute task completion. In any event, the close connection between linguistic and nonlinguistic goals in tasks in general suggest that linguistic

performance and task completion should interact in important ways. Yet, Long (2015) and Pallotti (2009) maintain that we do not fully understand how 'task completion' affects language learning. Kuiken *et al.* (2010) provide initial evidence that communicative adequacy in a task affects performance in classroom-based TBLT. Yet, in general, the relationship between linguistic performance and the completion of a communicative goal in tasks is poorly understood.

There are at least two predictions stemming from the research on how task completion and task performance (e.g., CAF) interact. First, learners focused on completing a task may produce language that is linguistically complex and accurate. Long (1989) hypothesizes that closed tasks (i.e., tasks where a possible answer or conclusion is derived) place communicative burdens on learners such that they strive to make themselves understood, resulting in greater linguistic complexity and accuracy; however, the empirical results are mixed. Manheimer's (1995) study revealed that closed tasks elicit more complexity, while Brown (1991) concluded that learners produced more complexity when engaged in an open task (an 'interpretive task') than in a closed task (a decision-making task). Tong-Fredericks' (1984) participants produced more complexity and accuracy during open tasks (i.e., role-play, interaction task) than in a closed task (a problem-solving task). Studies looking at the effect of task type on negotiation of meaning indicate that closed tasks generate more negotiation (Berwick, 1990; Newton, 1991; Pellettieri, 2000).

Second, when learners focus on completing a task's communicative objectives, their linguistic performance may be affected variably. One component may be robust (e.g., high complexity) whereas another may not (e.g., low accuracy). Skehan (2014) suggests that when learners are in the middle of a task and are focused on achieving a communicative goal, they are unlikely to monitor their output. Li (2014) reports that when learners focus on accuracy they also produce more syntactic complexity. Adams and Nik (2015) note that, compared to face-to-face (FTF) communication, learners engaged in text chat generate less syntactic complexity but higher levels of accuracy.

There is reason to suspect that the effect of task completion on task performance is likely to be strong in tasks that entail a high degree of authentic contextualization, such as CALL tasks involving a virtual world. According to Mroz (2014) a distinguishing feature of virtual contexts is that they promote learner agency (e.g., autonomy) and immersion in the L2. This combination of features makes 3D environments "approximate naturalistic L2 learning" (Mroz, 2014, p. 334). These environments approximate naturalistic environments in the sense that communication is goal (i.e., non-linguistically) oriented, be that for mundane purposes (e.g., buying a plane ticket) or for something more complicated (e.g., resolving a dispute). Lafford (2004), who comments on the developmental advantages of immersion settings, asserts that an emphasis on communicating ideas may cause learners "not to focus as much on the development of their L2 lexicon and structures" (p. 217) as they might in other contexts of learning. This suggests that tasks –which focuses learners on meaningful, situated language use– set in a virtual

world may encourage students to prioritize task completion. If so, a CALL-based task may provide important insights into whether and the extent to which L2 production (e.g., CAF) interacts with tasks' nonlinguistic, communicative objectives. Thus, the research question for the present study is:

Does the extent to which learners complete a CALL-based task in an immersive 3D environment affect the nature of their linguistic performance in the task in terms of complexity and accuracy? (These learners used an apostrophe in the place of a written accent mark).

Method

Participants. A total of 66 learners of Spanish enrolled in three advanced-level (i.e., third year) Spanish classes at a medium-sized university in the United States participated in the study. Commonly, learners at this level range from intermediate-low to intermediate-high on the American Council on the Teaching of Foreign Languages speaking proficiency scale (i.e., B1 to B2 in the Common European Framework). All participants had met or exceeded the learning outcomes from the previous course, a fourth-semester Spanish course. The classes were traditional, FTF courses employing group activities as well as a variety of multimedia activities (e.g., watching videos, Internet exploration/research). The focus of the courses was on increasing proficiency and accuracy in learner production. While the courses required writing, speaking, and inductive/exploration activities, they did not entail any chat or

instant messaging beyond the study. All students provided their informed consent to participate in the study.

Tasks. Learners participated in two tasks designed by the present author, which were integrated into two lessons lasting two class periods of 1.5 hours for a total of 3 hours and which were carried out in a laboratory equipped with Mac laptops. Each task contained a 3D exploration segment (authored in the Unity game development tool http://unity3d.com/unity/)[2] and a subsequent text chat segment, which occurred in a local area network via iChat, a (near) synchronous conference application.[3]

In the 3D exploration segment of each task, learners were first-person characters (FPCs) on a 3D island where they explored and collected clues by interacting with non-participatory characters (NPCs) –3D representations of humans– and objects (e.g., notes, letters). The 3D island used in this project was realistic, contained vegetation, and resembled a tropical island. The leaves of trees moved, and birds flew overhead, producing bird calls and other sounds. The space used for this study was centered around an ocean inlet whose waves undulated. Huts and wooden homes were seeded in the 3D island. The huts lined the shoreline, and some of these structures had outdoor furniture such as patio chairs near them. Between two of the huts there was a space where an outdoor fire pit had been used. The wooden homes were located a bit away from the shoreline, but still within view of the huts. Participants could freely and easily move between the huts and wooden homes. Each hut and home had a label indicating the name of the

resident that inhabited it. Huts had a wooden placard on an outside wall near the door, while homes had a mailbox with the NPC's first name printed on it near the front of the property line. Inside, each hut or home contained furniture such as a bed, a chest of drawers, paintings on the wall, rugs on the floor, etc.

By using the arrow keys, participants roamed within the 3D environment and freely chose which NPC or object to approach and how often. There were seven NPCs in all: two females and five males. Each was strategically seeded in the 3D world to encourage exploration. Some were standing near or inside their dwelling; one NPC was doing calisthenics near her home. All NPCs were created to be life-like and proportional to the 3D island.

The first task asked learners to find clues to solve a missing-persons case while the second, unrelated task required learners to solve a murder mystery.[4] When learners approached a NPC, three possible questions written in Spanish appeared on the screen. By clicking on one of the three textboxes, participants received a written answer – again, in Spanish. For the missing-persons task, the questions and answers were written to provide clues to determine the whereabouts of the missing person, while the ones for the murder mystery were written to provide details about each NPC's alibi on the day of the murder. For example, for the missing-persons task, Ana provided the following clue, as seen in Table 1.

Table 1
Example of a clue for the missing-persons task

| Questions that appear in textboxes	Answers that appear in textboxes
¿Dónde está Angela? 'Where is Angela?'	*No sé. Creo que está en la playa.* 'I don't know. I think she's at the beach.'
¿De qué hablaron? 'What did they talk about?'	*Me pidió el dinero que necesitaba para algo. No sé para qué era.* 'She asked me for the money that she needed for something. I don't know what it was for.'
¿Por qué? 'Why?'	*Me dijo que alguien le había robado todo el dinero.* 'She told me someone had robbed all her money.'

An example of a clue for the NPC named Víctor in the murder-mystery is shown below in Table 2.

Table 2
Example of a clue for the murder-mystery task

Questions that appear in textboxes	Answers that appear in textboxes
¿Qué hiciste durante la fiesta? 'What did you do during the party?'	*Pues, hablé con mis vecinos, tomé vino, y comí la comida de doña Ana.* 'Well, I spoke with my neighbors, drank wine, and ate doña Ana's food.'
¿Viste algo raro en la fiesta? 'Did you see anything strange at the party?'	*Pues, al pensarlo bien, Tito se fue de la fiesta, pero no sé la razón.* 'Well, now that I think about it, Tito left the party, but I don't know why.'
¿Regresó? 'Did he return?'	*Sí, regresó. Creo que Tito regresó después de una hora, más o menos.* 'Yes, he returned. I believe that Tito returned after an hour, more or less.'

When learners approached an object, a written message in Spanish containing information to read (e.g., a learner could approach a diary to see an entry) appeared.[5]

As with the NPCs, participants explored the 3D island collecting clues in any order.
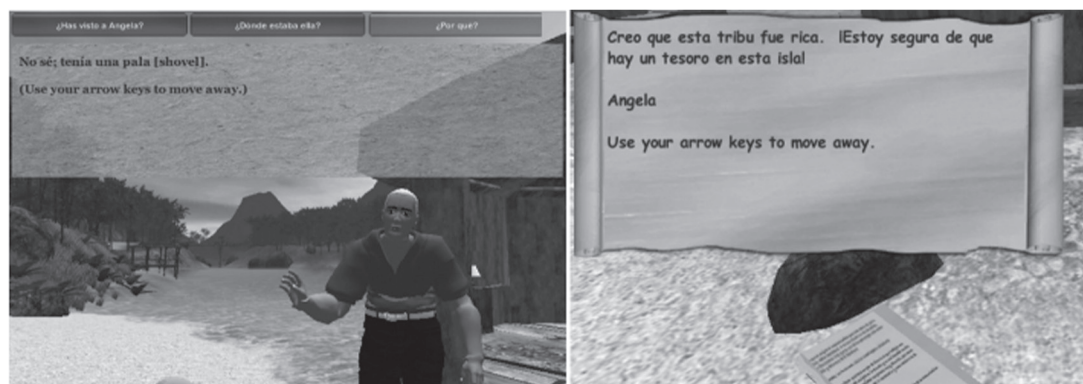
Figure 1. *Screenshot of 3D island.*

The 3D segment lasted 10 minutes, after which the learners closed the application.[6]

Next, immediately after the 3D exploration segment, students were paired up into randomly assigned dyads to complete the text chat segment, which lasted 25 minutes. Participants used the same computers in the same computer lab where they had been for the 3D segment. The researcher ensured that dyads were physically separated from each other in the laboratory so as to not make eye contact or so that they could not hear each other. Additionally, the researcher did not allow dyads to talk or make gestures to other participants during the text chat segment. Each dyad was to come to a consensus relating to the relevant crime by chatting in Spanish about the clues they had collected. For the missing-persons task, dyads were to determine the reason(s) for the person's disappearance; for the murder-mystery, dyads had to determine the reason(s) for the murder. To communicate these ideas in Spanish, one can use an embedded clause to express one's belief about an NPC's actions (e.g., *Creo que Juan se fue a otra isla con Ana* 'I think that Juan went to another island with Ana') and use the preterit and imperfect to narrate in the past (e.g., *Juan le escribió una carta de amor a Ana antes de irse* 'Juan wrote Ana a love letter before leaving'), both of which are taught in courses prior to the one in which these participants were enrolled and which are used regularly in the third-year Spanish courses at this institution. Indeed, the kind and level of linguistic-communicative proficiency required by the task and the kind and level of linguistic-communicative proficiency available to the individual learners for all intents and purposes were the same.

Analysis

A mixed-method analysis was employed, combining quantitative and qualitative perspectives (Greene and Caracelli, 1997). The quantitative analysis employed a stepwise regression analysis to examine the extent to which the production of four measures of linguistic complexity and accuracy in the chat segments predicted task completion (see *Dataset: Chatscripts (Independent Variables)* below). To provide a contextualized perspective, the qualitative analysis constituted an in-depth description of two (representative) dyads' production of linguistic complexity and a discussion of the extent to which they met the tasks' communicative objectives (see *Qualitative Analysis* below).

Dataset: Task Completion (Dependent Variable). To determine whether learners had met the tasks' communicative objectives, two judges independently assessed the chatscripts and awarded scores of 0, 1, or 2. Both judges were researchers in SLA and native speakers of Spanish. Prior to scoring the dataset, they attended a training session to practice assessing task completion with mock chatscript data. A score of 0 was awarded a chatscript in which the dyad did not complete the task, e.g., the dyad chatted about several possible criminals for the murder but never came to a consensus. A score of 1 was awarded when the dyad reached a consensus without producing a correct answer, e.g., the dyad decided that Víctor was the murderer when it really was Tito. A score of 2 was given when the dyad provided the correct answer, e.g., the dyad correctly determined that Tito was the

murderer. Since the dependent-variable scale was close ended and ordinal (similar to a Likert scale), inter-rater reliability was calculated with a Cohen's kappa coefficient, showing that there was almost perfect correlation between the judges' scores [к = .88]. To account for possible training-session effects (e.g., the potential for non-independence of raters), the researcher also calculated a Gwet's AC_1, which also showed the judges scores were highly correlated (Gwet's AC_1 = 0.88, SE = 0.35). For the regression analysis, the researcher summed the scores of the two rates to calculate a participant's task-completion score. Both inter-rater reliability analyses and the regression analysis were conducted with the R statistical computing package (R Core Team, 2017).

Dataset: Chatscripts (Independent Variables). Each dyad's iChat transcript was archived to a text document and collated by participant, allowing the measurement of various aspects of a learner's linguistic complexity and accuracy. The corpus totaled 27,315 words. This study measured complexity and accuracy but not fluency; three variables were calculated from the chatscripts to measure linguistic complexity and one variable was used to measure accuracy, all common to TBLT research. The first two response variables represent measures of lexical complexity (cf. Ellis, 2003). The third variable represents a measure of structural complexity, and the fourth linguistic accuracy.

Table 3
Response variables

Learner type token ratio (TTR) of a learner's production	The ratio of unique words to total words produced in the SCMC segment. The higher the ratio, the more unique words a learner produced.
Learner lexical density ratio of a learner's production	The ratio of unique main parts-of-speech (i.e., nouns, verbs, adjectives, and adverbs) words to total main parts-of-speech words in input read. The higher the ratio, the more semantically dense a learner's production was.
Learner clauses per c-unit	The number of clauses in a c-unit, an utterance containing a single complete sentence, phrase, or word and that has a clear semantic and pragmatic meaning in the context in which it occurs. The c-unit is similar to the T-unit, although it is more appropriate for the elliptical nature of conversations and SCMC (cf. Skehan, 1996).
Learner percentage of error-free clauses	The percentage of clauses a learner produced that contained no grammatical or lexical errors. All errors in syntax, morphology, and lexical choice were considered (cf. Ellis, 2003).

The type token and lexical density ratios were calculated for each learner with Wordsmith Tools, a concordance software program (Smith, 2016). To partially account for the effects of highly repeated terms in the chatscripts, the researcher set the stop boundary in Wordsmith for both the TTR analysis and the lexical density analysis at 662, such that on average the last 20% of each chatscript (827.7 * 0.80 = 662) was not overrepresented in the analysis. In any event, a particular participant's TTR was calculated by dividing the unique words s/he produced by the total number of words s/he produced in the SCMC portion of the experiment. The error-free clauses were derived by counting for each learner the frequency of clauses containing no errors divided by the total number of clauses the learner produced, which was calculated by summing the number of independent and dependent clauses per learner. For example, if a learner produced a statement such as *creo que Angela *está *el culpable* 'I think Angela is the guilty one', s/he would be counted as having one error-free clause (i.e., *creo*) and one erred clause (i.e., *que Angela *está *el culpable*). An independent clause represented the first clause of a c-unit. A dependent clause was headed by either a subordinating conjunction such as *que* 'that' or a coordinate conjunction such as *y* 'and' or *o* 'or'.

An inter-rater reliability analysis was employed to check the construct validity of the researcher's reading of erred clauses and number of clauses per c-unit since identifying both constructs requires a certain degree of judgment. Specifically, an experienced researcher was presented with a random sample of 100 segments from the corpus. She was to both count the number of clauses per c-unit and errorless clauses. Her two sets of scores were compared with those of the researcher with a Pearson correlation, as each of the 4 datasets ($n = 100$) was on an interval scale. Concerning the count of errorless clauses, the correlation between the two researchers was significant [r ($df = 99$) = .92, $p = .01$]. Regarding the count of c-units, the correlation between the two researchers was also significant [r ($df = 99$) = .91, $p = .01$].

Regression Analysis. A (both-directions) stepwise regression analysis was employed to determine which combination of independent variables best predicted the dependent variable. A regression analysis also indicates whether independent variables are correlated (i.e., have a positive coefficient) or disassociated (i.e., have a negative coefficient) with the dependent variable. In other words, the analysis, whose results are presented below, will indicate whether a combination of TTR, lexical density ratio, clauses per c-unit, and percentage of error-free clauses (and whether as one of these scores increases the other decreases, or not) is significantly associated with task completion in this study.

Results and General Discussion

Quantitative Analysis. Regarding the quantitative analysis, the mean task completion score was 1.23 [$sd = .65$], indicating that roughly 61% of the dyads successfully identified the murderer/solved the missing-persons case. Regarding the linguistic complexity measures overall, over half of the words that dyads produced were considered unique words, with a mean TTR of 0.55 ($sd = 0.07$), while about 65% of the

total content words dyads produced were unique content words (mean lexical density ratio = 0.65; *sd* = 0.07). These scores of lexical uniqueness are quite high given the average amount of words per chat (827.7 words), indicating that chatscripts contained a good amount of lexical diversity and density. With respect to the learners' accuracy, dyads produced on average almost 25 error-free clauses per chatscript (mean error-free clauses = 24.86; *sd* = 8.91). Thus, given that dyads averaged 74.9 clauses per chat, approximately 1/3 of their clauses were error free, which is probably not unusual for third-year learners. Finally, the learners generated almost two clauses for every c-unit (mean clauses per c-unit = 1.8; *sd* = 0.45).

The regression analysis [$F(2,63)$ = 5.01, p = 0.01, R^2 = 0.11] indicates that, of the 4 independent variables, two *combine* significantly to predict completion of the task: percentage of error-free clauses (standardized β = -0.01,

p = .155, SE = 0.01) and clauses per c-unit (standardized β = 0.48, p = .006, SE = 0.17). The regression analysis indicated that neither lexical-complexity measure (i.e., TTR or lexical-density ratio) predicted task completion.

Since clauses per c-unit was the only predictor whose coefficient was significant, it is important to discuss how both predictor variables interact to predict task completion. Even though the error-free clauses was deemed an important predictor of task completion, it does not predict task completion in an *entirely* linear fashion, as its standardized beta is, for all intents and purposes, zero. Clauses per c-unit, however, does predict task completion in a linear fashion: it seems that, as learners produce more clauses per c-unit, their chances of task completion increase. Figures 2 and 3 provide a more detailed analysis of the two predictor variables vis-à-vis task completion, breaking down the contribution of each predictor variable by task-completion score.
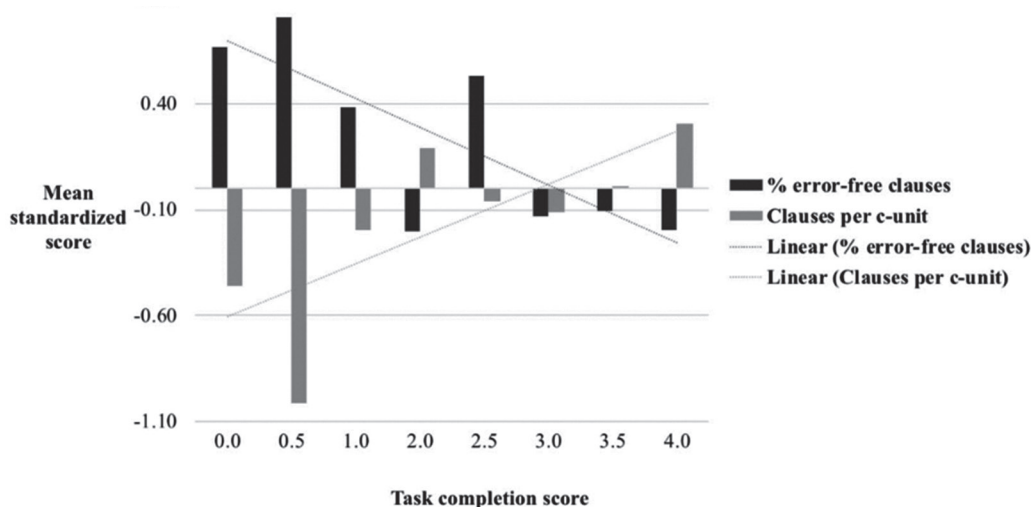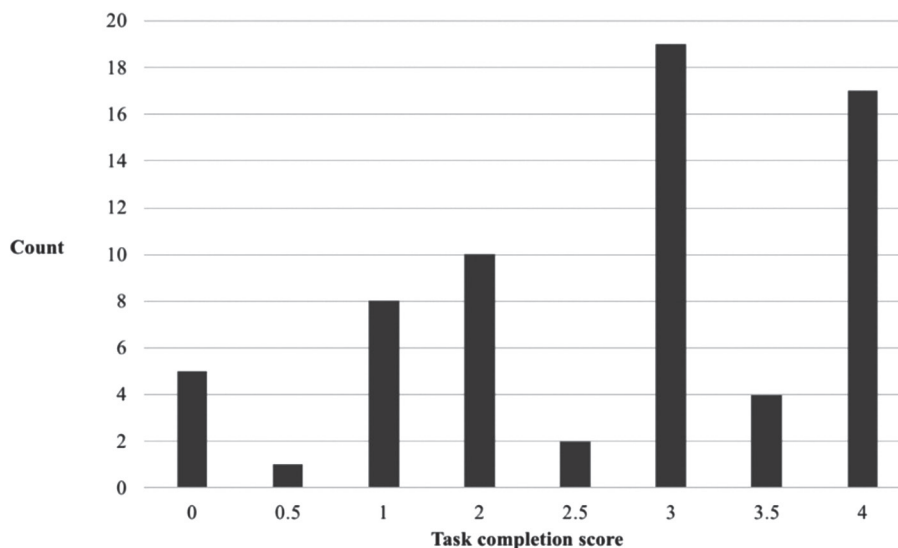

Figure 2. *Standardized predictor variable scores and trend line by task-completion score.*

Figure 3. *Participant count by task-completion score.*

Considering the two trend lines in Figure 2, this analysis suggests that, as learners' task completion scores increased, the following interaction occurred: learners produced fewer error-free clauses and more clauses per c-unit. In other words, task completion was associated with more syntactic complexity but also more errors. Still, the analysis indicates that this trend held mostly for learners who had low or high task-completion scores. Learners who scored between 0.0 and 1.0 exhibited relatively little syntactic complexity and few errors. Learners who scored the maximum of 4.0, conversely, exhibited a relatively good amount of syntactic complexity and errors. A total of 47% (31/66) of the learners constituted these two extremes. The remaining learners did not exhibit this complex interaction between complexity, errors and task completion. This fine-grained analysis explains two aspects of the statistical analysis: (1)

how and why percentage of error-free clauses and clauses per c-unit predict task completion; (2) the fact that, although the model was significant, the amount of explained variation (i.e., R^2) was decidedly low. Overall, those dyads that met the tasks' communicative objectives were those that produced more clauses and at the same time more errors than their peers. Conversely, dyads less likely to complete the task were those that generated few clauses and yet relatively few errors.

Qualitative Analysis. Below, the researcher includes and discusses portions of two chatscripts, one that is highly representative of the regression model and one that does not represent the model well at all. Each was selected numerically, based on the regression analysis. Based on a z-score analysis of the participants' clauses per c-unit and percentage of error-free clauses, the chatscript highly representative of the

regression model contains a markedly high number of clauses per c-unit and few error-free clauses. Conversely, the chatscript poorly representative of the regression model contains a markedly low number of clauses per c-unit and several error-free clauses.

In the first chatscript (see Appendix 1) learners KAB and NMC are near the end of their chat, piecing together the clues they gathered in the exploration phase on the 3D island. KAB suggests that it is odd that Juan saw the missing person (Angela) but no one else has seen her for four days (*[creo] [que es un poco raro] [que el la *vista en el mismo dia] [cuando el resto de la gente la *vista *hasta cuatro dias.*] 'I think that it is a bit odd that he saw her the same day when the rest of the people saw her four days ago') while NMC writes that someone he interviewed saw Angela with a shovel ([**encuentre muchas cosas y gente] [*quien *dice la *vista con un *patlil *"shovel?*] 'I found many things and people who said that they saw her with a shovel'). In the turn marked in bold font, the dyad completes the task; KAB agrees with NMC about the shovel and informs him that she saw a note in Angela's hut about going to the other side of the island ([*Si, yo tambien.] [Tambien *ve una nota en la casa de angela] [que ella escribio] [que dice] ["voy al otro lado de la isla] [y llevo un shovel"] [o algo similar]* 'Yes, me too. I also saw a note in Angela's house that she wrote that said, "I'm going to the other side of the island and I'm taking a shovel" or something like that').

Concerning the linguistic complexity of this chatscript, KAB and NMC generate numerous clauses per c-unit. Indeed, in the turn in which KAB completes the task, there are some seven

clauses. Accompanying the large number of clauses per c-unit are numerous errors, including lexical errors (*shovel*, *treasure*, but also *vista* 'view'), grammatical structure errors (*hasta cuatro días* 'until four days' when the learner probably meant *hace cuatro días* 'four days ago'), morphology (*hablo* 'I speak' when context dictates that the learner meant *hablé* 'I spoke'), and agreement (*todo su posesiones* 'all his possessions' when correct adjective agreement would have generated *todas sus posesiones*). Overall, the combination of a large number of clauses per c-unit and numerous errors coincides with the fact that the dyads met the tasks' communicative objective and correctly determined where the missing person was.

Chatscript 2 (see Appendix 2) is quite different; it does not reflect the model. These learners (TB and JK) produce few clauses per c-unit and few errors. It is important to note that TB and JK do not complete the task; they do not come to a consensus about the murder mystery. In two consecutive turns, JK suggests that Pedro and Victor are suspects, [*creo] [que Pedro es un *suspecto*] 'I believe that Pedro is a suspect' and [*y victor tambie'n*] and 'Victor too'. The dyad continues along with this argument, suggesting that Victor might be the murderer because he left the party for an hour or more [*si, y victor *fue' * la fiesta *para *un hora ma's o menos.*] 'yes, and victor left the party for an hour more or less', but they never decide that Victor is the culprit. Near the end of their chat, they blame Pedro, stating that he had a motive [*pienso] [que es la *falta de pedro tambie'n] [porque e'l tiene los *motives porque e'l dijo] [que nadie *cae *bein *consigo*] 'I think that it is Pedro's fault

too because he has a motive because he said that he doesn't get along with anyone'. In their next exchange, they mention Tito as a possible murderer writing, [*tito *era *a la fiesta *solomenta *parta *triente minutos*] 'tito was at the party for only thirty minutes'. TB and JK offer various suggestions to solve the murder mystery, but they never come to a consensus. They did not meet the communicative objective of identifying the murderer.

A close examination of this dyad's chatscript portion reveals a relatively small number of clauses per c-unit; indeed, almost all of the c-units have only one or two clauses, in contrast to the previous chatscript portion where KAB and NMC produced on average six clauses per c-unit. This dyad's longest c-unit contains four c-clauses. In addition to few clauses per c-unit, this dyad produced relatively few errors. In this segment, TB and JK produced 11 clauses that were entirely error-free. This dyad did not meet the task's communicative objective, and their production was characterized by the combination of few clauses per c-unit and relatively few errors.

Summary and Conclusions

Recent theoretical discussions on how to evaluate the efficacy and outcomes of TBLT challenge researchers to consider not just learner performance, such as measures of CAF. Even though SLA research has focused on understanding what happens to L2 production when learners are focused on a nonlinguistic, communicative goal, it has largely ignored the effects of both production on task

completion and task completion on production. Yet, SLA theory clearly posits that task performance is affected in important ways by the attention that learners place on attaining a task's communicative objective (Long, 2015; Pallotti, 2009; Skehan, 2014). The study reported here provides empirical support for Pallotti's (2009) conjecture that task performance should be considered alongside task completion, since the two constructs were shown to interact significantly (e.g., deficient task performance can nonetheless lead to successful task completion, and vice versa). The CALL-based quantitative analysis reported in this study reveals that third-year learners of Spanish who met the task's communicative goal produced discourse containing a large number of clauses per c-unit but numerous errors. Conversely, learners who did not meet a task's goal generated language containing a small number of clauses per c-unit but few errors. The qualitative analysis also provides empirical support for the symbiotic relationship between task completion and task performance, especially as it relates to learners of Spanish participating in tasks. For example, one dyad who completed the task produced discourse containing a large number of clauses per c-unit but also numerous errors, whereas another dyad who did not complete the task generated discourse containing few clauses per c-unit and more error-free clauses. It appears that when learners focus on solving a task (i.e., on communication in the L2) when engaged in text chat, instead of on form, they produce numerous propositions; yet, such production contains low levels of accuracy. It is important, nonetheless,

to note that this pattern held for approximately half of the learners, who were either very unsuccessful at task-completion or very successful. It is unclear why learners who had 'average' task-completion scores did not exhibit the relationship identified here. It may be that, amongst such learners, their attention shifted between the linguistic and non-linguistic goals of the task. Future research will need to explore this issue further.

Admittedly, virtual environments may focus learners on communicative outcomes more so than classroom-based tasks. Yet, the present author argues that virtual environments are a particularly good environment for studying the effects of communicative goals on L2 production, since virtual environments are immersive, and they promote agency (Taguchi and Sykes, 2013). When designed with task-based features, virtual environments can encourage learners to focus on meeting (non-linguistic) communicative objectives. This was the case in the present project: learners were tasked with identifying the murderer in the murder-mystery task and determining the location of the missing person in the missing-persons case.[7] Compare a reading task and a task set in a virtual environment. In a reading task, the main stimuli for learners are letters on a page. Language form (e.g., letters, punctuation) is front and center; however, focusing on meaning requires that the learner 'imagine' the context, which is challenging. By comparison, in a naturalistic 3D world, the main stimuli for focusing on meaning are not linguistic but rather visual (e.g., objects, people, places) and spatial (i.e., learners' movement in the world) stimuli. Regardless of the nature of the task, since learners are focused on communicative objective(s) and since they construct a mental representation of a 'situation' with visual and spatial cues, virtual environments can simulate naturalistic linguistic settings. Remaining cognizant of the non-linguistic aspects of the task (e.g., Who committed the crime?; Where is the missing person?) is almost assured. Regarding agency, in the virtual environment, the learner becomes in charge of the flow of information: if the learner does not interact with objects and avatars and does not process language form for meaning, s/he gathers no information to complete the task. In the present project, these autonomous exploratory behaviors yielded linguistic information to solve the two tasks.

Despite the study's large sample size, there are limitations to its generalizability. While all learners were enrolled in third-year courses, the variability in the participants' overall proficiency in Spanish is not accounted for. A proficiency level predictor variable in the analysis could reveal whether proficiency level in speaking and/or reading mitigates task completion and production (cf. Collentine [2015] for a study on how reading abilities affect learners' production in tasks). Additionally, even though this study employed measures representing linguistic complexity and accuracy that are common to task-based research; the metrics surely mask more subtle effects of task completion on production. Future research could include more fine-grained measures. Finally, regression analysis is considered a first step in establishing causality

(Tabachnick and Fidell, 2001), and future research could provide more confirmatory analyses (e.g., structural equation modeling). The small effect size (R^2 = 0.11) indicates that, even though the two variables identified here (clausal complexity and accuracy) significantly predict task completion, there exist other factors accounting for task completion that the present analysis could not uncover. This suggests that task completion is a function of multiple factors, which may relate to individual factors (e.g., autonomy, interest in the task, proficiency) along with the linguistic performance factors identified here.

In light of these considerations of the interpretation of the data, the author is not hypothesizing that there is a direct causal relationship between linguistic performance and task completion. Rather, the relationship is likely to be an indirect one of cause and effect, possibly mediated by other factors: as learners generate more language (i.e., more clauses per c-unit) and attend to meaning over form (i.e., thus, leading to more errors), they are more likely to reason their way to a conclusion that considers all the facts of the tasks and their implications for the goal established by the instructor.

What advice do these data offer teachers and materials developers of CALL for learners of Spanish? When learners are engaged in a closed-text chat task such as the one described here, teachers should not use an external measure such as CAF to determine success; rather, teachers should determine success on whether learners meet the task's communicative objectives. For example, in a missing persons case task such as the one in this study, instead of looking for specific syntax (e.g., noun clauses) or a set of grammatical constructs (e.g., subjunctive or indicative) in learner production, teachers should look for whether learners meet the task's goals, i.e., whether they draw a conclusion about the case based on evidence. Materials designers should create tasks that emphasize this focus on meeting communicative objectives as well. Those could include determining reasons for and against some architectural addition (i.e., a new building) in a town, deciding what tourist attractions to visit and why, coming to a consensus as to a celebrity guest list for a party, etc. In addition, teachers should expect their learners' production while engaged in tasks to vary in terms of CAF.

Notes

1.   The present study forms part of a larger study on 3D worlds and tasks. Collentine (2013) reports on the materials and portions of the analysis included in the present study, specifically, the study's predictor variables. However, the response variable – namely, task completion – is not reported elsewhere. Additionally, neither the regression analyses entailing task completion, the interrater reliability analysis, nor the qualitative analyses are reported elsewhere.

2.   Unity allows applications to be built for Windows, Macs, or Linux. For this study, the application was built for Mac since the laboratory available for the data collection contained only Mac computers.

3.   Any other instant messaging platform, e.g., Moodle Chatroom, could have also been used for the experiment.

4. No special experience or training in criminal investigation was necessary for participants. The skills necessary to solve the tasks were similar in cognitive complexity to those needed to play the board game *Clue*, a game whose intended audience is children ages 8 and older. Additionally, Pica, Kanagy, and Falodun (1993) specifically recommend that TBLT engage learners in mysteries where they must search and assess clues with respect to some question (e.g., a crime). Such activities are a type of jigsaw task, one of the fundamental TBLT tasks types.

5. To familiarize the learners with both the 3D technology and iChat, the day before the experiment the learners navigated a sample 3D world (not employed in the present analyses) containing examples/instances of the technologies described here.

6. The researcher emphasized that participants should take advantage of the affordances of the 3D world to solve the task.

7. The reader is reminded that, while very basic critical thinking skills were needed to solve the missing-persons case and the murder case, no special training or experience was needed to successfully complete the tasks. Additionally, Pica *et al.* (1993) identify tasks requiring a solution to mystery based on students gathering clues as a valid type of TBLT jigsaw activity.

Bibliography

Adams, R. & Nik, N. (2015). Prior knowledge and second language task production in text chat. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT* (51-78). Amsterdam/Philadelphia: John Benjamins.

Berwick, R. (1990). *Task variation and repair in English as a foreign language*. Kobe, Japan: Kobe University of Commerce: Institute of Economic Research.

Brown, R. (1991). Group work, task difference, and second language acquisition. *Applied Linguistics*, 21, pp. 1-12.

Collentine, K. (2015). The effect of reading on second-language learners' production in tasks. *Hispania*, *99(1),* pp. 51-65.

Collentine, K. (2013). Using tracking technologies to study the effects of linguistic complexity in CALL input and SCMC output. In P. Hubbard, M. Schulze, and B. Smith (Eds.), *Learner-computer interaction in language education: A Festschrift in honor of Robert Fischer*. (pp. 46-65). San Marcos, TX: CALICO.

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

Greene, J. & Caracelli, V. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J. Greene, & V. Caracelli (Eds.), Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (5-18). San Francisco, CA: Jossey-Bass.

Kuiken, F., Vedder, I. & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (81-100). Roma: Eurosla.

Lafford, B. (2004). The effect of the context of learning on the use of communication strategies by learners

of Spanish as a second language. *Studies in Second Language Acquisition*, 26, pp. 201-225.

Li, Q. (2014). Get it right in the end: The effects of post-task transcribing on learners. In P. Skehan, (Ed.), Processing perspectives on task performance (129-154). Amsterdam/Philadelphia: John Benjamins.

Long, M. (2015). *Second language acquisition and task-based language teaching*. West Sussex, United Kingdom: Wiley and Sons.

Long, M. (1989). *Task, group, and task-group interactions*. University of Hawaii Working Papers in ESL.

Manheimer, R. (1995). Close the task: Improve the discourse. Paper given at *Annual Conference of American Association of Applied Linguists*. Long Beach, CA.

Mroz, A. (2014). 21st century virtual language learning environments (VLLEs). *Language and Linguistics Compass, (8)*8, 330-343.

Newton, J. (1991). Negotiation: Negotiating for what? Paper given at *SEA-MEO conference on language acquisition and the second/foreign language classroom*. RELC, Singapore.

Pallotti, G. (2009). CAF: Defining, refining, and differentiating constructs. *Applied Linguistics*, 30, 590-601.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communicative tasks for second language instruction. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (9-54). Clevedon, England: Multilingual Matters.

Pelletieri, J (2000). Negotiation in cyberspace. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching* (59-86). Cambridge: Cambridge University Press.

R Core Team (2017). *R: A language and environment for statistical computing*. R Vienna, Austria: Foundation for Statistical Computing. URL https://www.R-project.org/.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, pp. 27-57.

Skehan, P (2014). Limited attentional resources, second language performance, and task-based pedagogy. In P. Skehan (Eds.), *Processing perspectives on task performance* (211-260). Amsterdam/Philadelphia: John Benjamins.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*, pp. 38-62.

Scott, M. (2016). *WordSmith Tools version 7*. Stroud: Lexical Analysis Software. URL: https://lexically.net/wordsmith/

Tabachnick, B. and Fidell, L. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn and Bacon.

Taguchi, N. and Sykes, J. (2013). *Technology in interlanguage pragmatics research and teaching*. Amsterdam/Philadelphia: John Benjamins.

Tong-Fredericks, C. (1984). Types of oral communication activities and the language they generate: A comparison. *System*, 12, pp. 133-134.

Appendix I

Chatscript 1. Highly representative of the regression model.

KAB: [si, estaba en su casa cuando la *vista.] [y si hable con juan tambien.] [creo] [que es un poco raro] [que el la *vista en el mismo dia] [cuando el resto de la gente la *vista *hasta cuatro dias.] [hablaste con nora?] [porque no *puedo]

NMC: [*a mi tampoco,] [no pude hablar con nora.] [si creo] [que es un poco raro] [que la otra gente la *vista hace cuatro dias,] [y juan hoy.] [*encuentre muchas cosas y gente] [*quien *dice la *vista con un *patlil *"shovel?"]

KAB: [Si, yo tambien.] [Tambien *ve una nota en la casa de angela] [que ella escribio] [que dice] ["voy al otro lado de la isla] [y llevo un shovel"] [o algo similar]

NMC: [si, *a mi tambien.] [no pude ir al otro lado de la isla.] [pudiste tu?]

KAB: [no, no pude tampoco.]   [asi no se] [porque ella necesita un *shovel.] [tienes ideas?]

NMC: [*hablo con donna ana] [y ella dice] [que hay *"treasure" en esta isla,] [creo,] [es posible] [que angela *fue a buscar el *"treasure"] [*por que la gente ha robado *ella de *todos *su posesiones y su dinero]

Legend:

- Clauses are segmented within brackets […] .
- Errors are marked with an asterisk * .
- Bolded clauses indicate c-units where dyad completed the task.

Appendix II

Chatscript 2. Poorly representative of the regression model.

TB: [que encontraste?]

JK: [ma's informacio'n][1]

TB: [yo tambie'n]

JK: [creo] [que Pedro es un *suspecto]

JK: [y victor tambie'n]

TB: [yo tambie'n,] [porque no *sali' a la fiesta?]

JK: [tambie'n]

TB: [*porque *pienso] [que era Victor?]

JK: [si, y victor *fue' * la fiesta *para *un hora ma's o menos.]

…

TB: [pienso] [que es la *falta de pedro tambie'n] [porque e'l tiene los *motives porque e'l dijo] [que nadie *cae *bein *consigo]

JK: [si..muy *suspicioso.]

TB: [tito *era *a la fiesta *solomenta *parta *triente minutos]

TB: [verdad?]

TB: [o no]

Legend:

- Clauses are segmented within brackets […] .
- Errors are marked with an asterisk * .
- Bolded clauses indicate c-units where dyad completed the task.