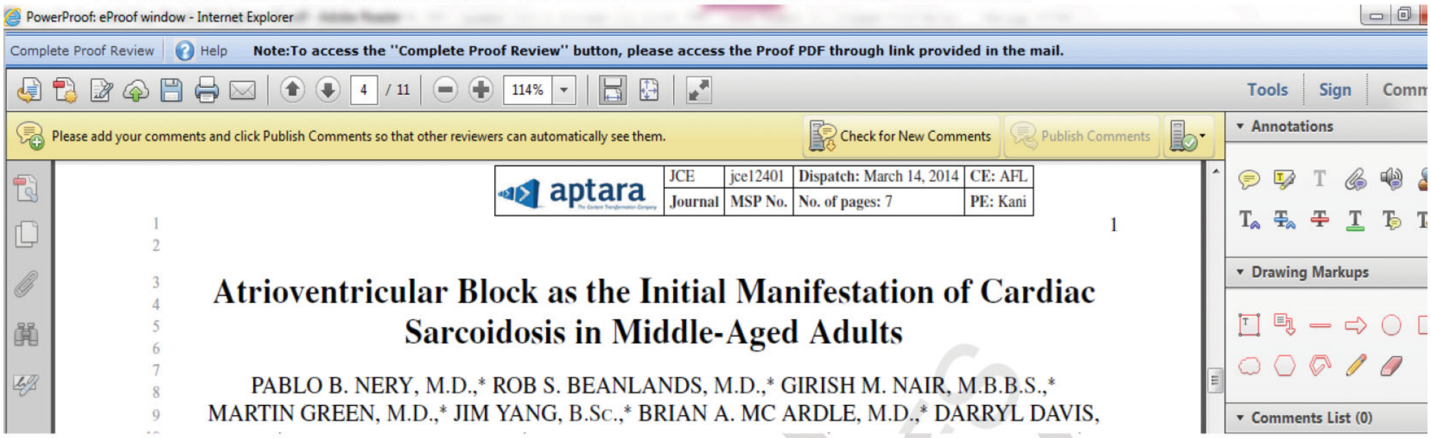
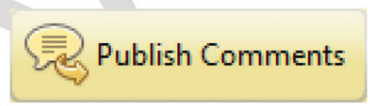
WILEY

Online Proofing System

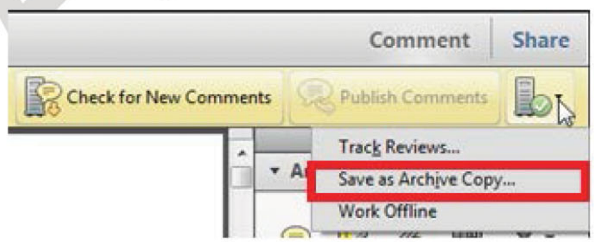
1. Corrections should be marked with the Adobe Annotation & Comment Tools below:



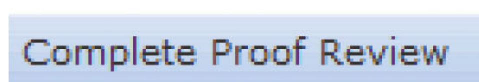
2. To save your proof corrections, click the 'Publish Comments' button. Publishing your comments saves the marked up version of your proof to a centralized location in Wiley's Online Proofing System. Corrections don't have to be marked in one sitting – you can publish corrections and log back in at a later time to add more.



3. When your proof review is complete we recommend you download a copy of your annotated proof for reference in any future correspondence concerning the article before publication. You can do this by clicking on the icon to the right of the 'Publish Comments' button and selecting 'Save as Archive Copy...'.



4. When your proof review is complete and you are ready to send corrections to the publisher click the 'Complete Proof Review' button that appears above the proof in your web browser window. Do not click the 'Complete Proof Review' button without replying to any author queries found on the last page of your proof. Incomplete proof reviews will cause a delay in publication. **Note: Once you click 'Complete Proof Review' you will not be able to mark any further comments or corrections**.



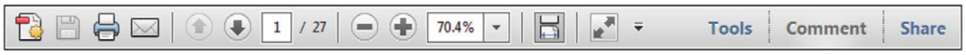
Firefox, Chrome, Safari Users

If your PDF article proof opens in any PDF viewer other than Adobe Reader or Adobe Acrobat, you will not be able to mark corrections and query responses, nor save them. To mark and save corrections, please follow these instructions to disable the built-in browser PDF viewers in Firefox, Chrome, and Safari so the PDF article proof opens in Adobe within a Firefox or Chrome browser window.

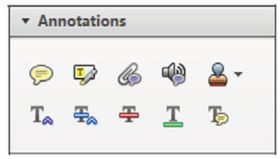
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 9.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:



This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

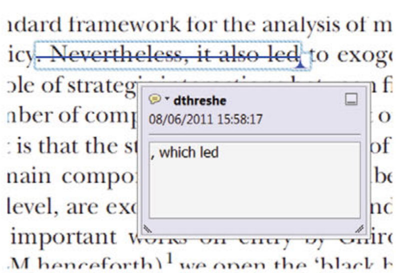


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

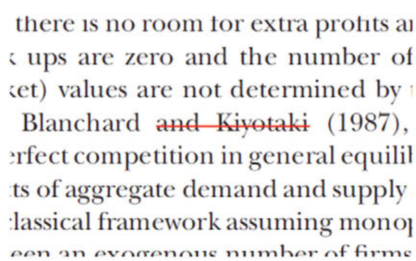


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

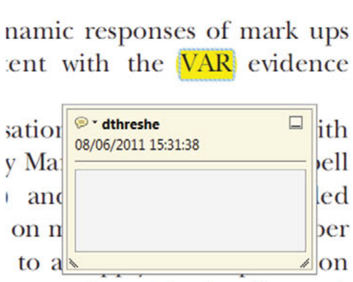


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

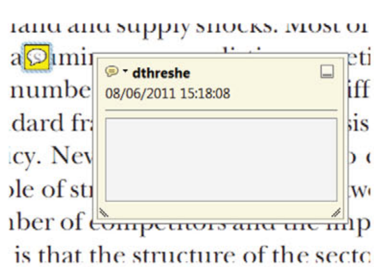


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.



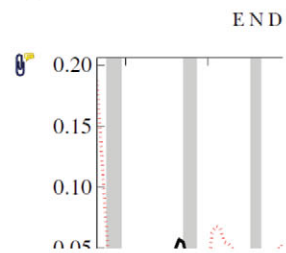
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate pace in the text.

How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

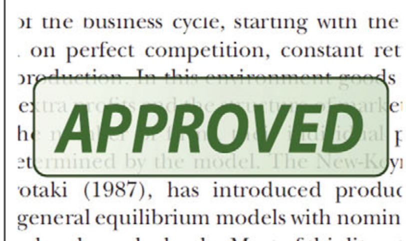


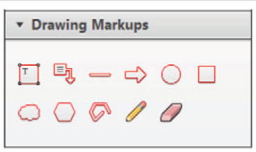
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it

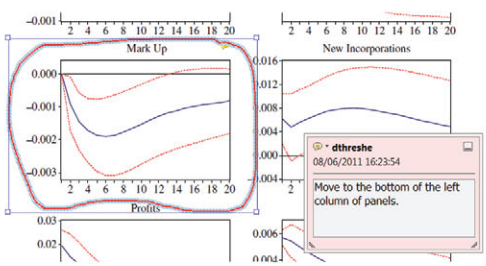
- Click on the Add stamp icon in the Annotations section.
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears).
- Click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).





7. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

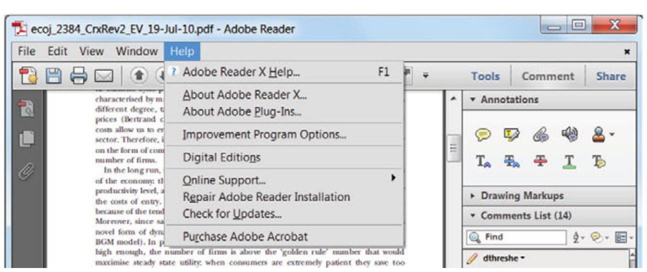
Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks..



How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.

For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



APTARA	MODL	modl12382	**Dispatch:** December 5, 2016	**CE:** XXX
	Journal	**MSP No.**	**No. of pages:** 15	**PE:** XXXX

L2 Speaking Development During Study Abroad: Fluency, Accuracy, Complexity, and Underlying Cognitive Factors

KAREN RUTH LEONARD[1] and CHRISTINE E. SHEA[2]

[1]*Department of English and Modern Languages, Olivet Nazarene University, One University Ave, Bourbonnais, IL, 60914 Email: krleonard@olivet.edu*

[2]*Department of Spanish and Portuguese, University of Iowa, 412 Phillips Hall, Iowa City, IA, 52243 Email: christine-shea@uiowa.edu*

We take a multidimensional perspective on the development of second language (L2) speaking ability and examine how changes in the underlying cognitive variables of linguistic knowledge and processing speed interact with complexity, fluency, and accuracy over the course of a 3-month Spanish study abroad session. Study abroad provides a unique learning context for evaluating changes in the underlying dimensions of L2 speaking because learners are fully immersed in the target language and have ample opportunity to implement, practice, and integrate newly gained skills. Participants were 39 native English speakers acquiring Spanish in Argentina. Results show that participants experienced significant gains across complexity, fluency, and accuracy. However, these gains were not evenly distributed across all dimensions or across all learners. Learners with higher levels of L2 linguistic knowledge and faster L2 processing speed prior to study abroad experienced greater gains in accuracy and syntactic and lexical complexity during study abroad.

Keywords: study abroad; fluency; accuracy; complexity; Spanish

IN SECOND LANGUAGE ACQUISITION (SLA), learning context plays a crucial role in target language development, influencing both the rate of acquisition and the final outcome (Sanz & Grey, 2015). To a great extent, the learning context determines the balance in type of input received by the learner (explicit vs. implicit), the type of interaction required of the learner (meaning-based vs. form-based), and also the aspects of the input that are required to carry out day-to-day activities. Recently, investigators have recognized that SLA during study abroad (SA) provides an opportunity to examine these interaction–type effects in a unique learning context, in which all of these factors related to context come together (Sanz, 2014) and potentially create opportunities for target language development distinct from those available in a classroom context. For language learners, SA provides an ideal opportunity to accelerate their learning process (Sanz, 2014) and, in particular, to make rapid improvement to their speaking ability. Being fully immersed in a language provides exceptional opportunities for students to interact in that language and "soak it up … without even trying" (Sanz, 2014, p. 1).

In the present study, we take a multidimensional perspective on second language (L2) speaking development and examine how the various linguistic, psycholinguistic, and cognitive correlates of complexity, accuracy, and fluency (CAF) change and interact in the development of L2 speaking proficiency over the course of a 3-month study abroad session. Learning to speak in an L2 is a highly complex skill that draws upon linguistic knowledge and processing abilities to produce a fluent, accurate, and complex verbal message that

The Modern Language Journal, 00, 00, (2016)
DOI: 10.1111/modl.12382
0026-7902/16/1–15 $1.50/0

can be understood by interlocutors in an easy and efficient manner.

By situating our learners in an SA context, we provide a more complete picture of how L2 speaking ability changes over time. While DeKeyser (2007) argues that SA does not always bring about sizeable linguistic gains, he acknowledges that classroom learning and SA are not necessarily in opposition, as the declarative and procedural knowledge gathered in more formal classroom contexts can serve as a basis for the development of automatized language use in the SA context, in particular because learners living in an environment where the language is spoken feel the pressure to communicate orally in real time and to do so in a way that facilitates rapid, effective communication.

As far as we are aware, the current study is one of the first to consider the three constructs of complexity, accuracy, and fluency, connect them to their cognitive, linguistic, and psycholinguistic underpinnings, and observe changes over a 3-month SA session. By examining speaking ability at this level of granularity, we gain considerable insight into what exactly changes—and how these changes are interrelated—when learners are immersed in the target language. Participants in the current study were 39 native English speakers studying abroad in Buenos Aires for at least one semester.[1] In the following section we define the underlying dimensions of speaking considered in this study and present the tasks used to analyze them.

Q2

SPEAKING AS A COMPLEX CONSTRUCT: COMPLEXITY, ACCURACY, AND FLUENCY

As Housen, Kuiken, and Vedder (2012) state, CAF "reflect[s] the major stages of change in the underlying L2 system" (p. 3). Changes in complexity, for example, involve the incorporation of new L2 elements, while changes in accuracy include increasingly target-like interlanguage. Changes in fluency, on the other hand, require the consolidation and proceduralization of L2 knowledge. These three dimensions have been shown to be independent constructs (e.g., Norris & Ortega, 2009) but at the same time intricately interrelated dimensions of L2 proficiency that may not always develop linearly or at the same pace (Spoelman & Verspoor, 2010).

Complexity can be understood as either cognitive complexity or linguistic complexity; that is, as a learning problem or as an inherent characteristic of the structures themselves (DeKeyser, 2016; Housen et al., 2012). In the present study, we consider changes in complexity of the language produced by participants, specifically, changes in lexical and grammatical complexity. In terms of the latter, little previous research has examined how syntactic complexity may change during SA, and the studies that have examined this have led to inconclusive results. Mora and Valls Ferrer (2012) found no significant changes in the mean number of clauses per AS-unit,[2] a measure of subordination. Serrano, Tragant, and Llanes (2012), in contrast, found a significant increase in the mean number of clauses per T-unit.[3]

In terms of lexis, researchers have employed what might most accurately be called *lexical variety* (or *lexical diversity*), using a type–token ratio to measure how many different words are present in participants' speech samples. In regard to changes in lexical variety during study abroad, results from previous studies have not revealed any significant changes (Mora & Valls Ferrer, 2012; Serrano et al., 2012). The present study includes a measure of lexical variety (VocD) as well as an additional measure (the Guiraud advanced index, which we call a measure of *lexical complexity*) that takes into account whether the words produced are of high or low frequency; the use of more precise, low-frequency vocabulary is considered a feature of higher levels of speaking proficiency (Swender, Conrad, & Vicars, 2012).

The second dimension of speaking that we consider is *accuracy*. Learner accuracy in speech refers to the ability to produce error-free speech (Housen & Kuiken, 2009); in the study abroad research, *errors* generally include both grammatical and lexical deviations from native-speaker norms. Previous research on changes in accuracy during SA is limited, but recent studies have found some evidence of gains in this area (Llanes & Muñoz, 2009; Mora & Valls Ferrer, 2012; Serrano et al., 2012).

The final dimension we address is *fluency*, which can be understood as separate from complexity and accuracy but nonetheless closely connected in the development of L2 proficiency. Fluency refers to the temporal characteristics of speech, including such aspects as pausing, speed (speech rate), and repair (how often speakers make false starts or self-corrections). Previous study abroad research has provided fairly consistent evidence of gains in fluency; most studies found an increase in speech rate, measured as syllables or words per second or minute (D'Amico, 2010; Lennon, 1990a; Llanes & Muñoz, 2009; Mora & Valls Ferrer, 2012; O'Brien, Segalowitz, Freed, & Collentine, 2007; Serrano et al., 2012; Towell, 2002), and mean length of run (D'Amico, 2010;

Mora & Valls Ferrer, 2012; O'Brien et al., 2007; Towell, 2002). Results are less conclusive for measures of repair and pausing.

We consider changes in fluency as reflecting changes in the underlying components related to automaticity. Automatic processes are those that require no or limited attention, effort, or control and can be the result of conscious practice. When automaticity involves a complex process, any particular component of this process may (or may not) be automatized and is a result of procedural knowledge. A clear consequence of a lack of automaticity is that the underlying subcomponents of fluency will take longer to be carried out, or require more concentration on the part of the speaker. A speaker will necessarily slow down when executing the components of speaking that are not fully automatized, leading to slower, more hesitant, and possibly less accurate speech.

In Levelt's (1989) influential speech production model, language production is the result of a highly automatized system in which speakers conceptualize the information to be conveyed, formulate the linguistic structure necessary to produce it, and subsequently articulate the message itself. In first language (L1) speech, these processes operate with a high degree of automaticity and largely in parallel (Levelt, 1989). In L2 speech production, however, there are various points where the automaticity characteristic of L1 speech potentially breaks down, leading to what ~~is~~ perceived as nonnative-like productions.

As dimensions of speaking proficiency, CAF reflects the automaticity of procedural knowledge that reveals in turn the automaticity of subcomponent skills that underlie speaking. In the present study, we address this by including tasks that measure linguistic knowledge and processing speed (De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012). Participants carried out two knowledge tests (untimed vocabulary and grammar tests), a timed picture-naming task (for lexical retrieval), and a timed sentence–picture verification task (measuring morphosyntactic processing). By using timed and untimed tasks, we gain a better picture of what participants know and can express when time is not an issue (i.e., knowledge that has not necessarily been automatized) and what participants can express when they are under time pressure and presumably need to rely upon more automatized skills and knowledge. Indeed, proficiency in speaking involves declarative knowledge and the ability to process knowledge quickly (De Jong et al., 2012), both of which contribute to CAF. To this end, the tasks we use tap into the formulation stage of Levelt's model, the stage at which speakers retrieve lemmas from the lexicon to formulate sentences. These tasks round out the picture we present regarding speaking as a componential skill that includes complexity, accuracy, and fluency and also includes the underlying psycholinguistic processes and linguistic knowledge that ~~allows~~ speakers to produce complex, accurate, and fluent speech.[4]

TABLE 1
Participants

Characteristic	Mean	*SD*
Age at pretest	20.6	1.6
Age of onset (learning Spanish)	12.8	4.6
Years of formal study of Spanish	6.0	2.4
Previous time abroad in Spanish-speaking countries (months)	1.0	1.7
Number of other languages studied	0.7	0.8

In the following sections, we first present the methods (tasks and analyses) employed to operationalize each CAF dimension, linguistic knowledge, and processing speed. Following this, we present the research questions guiding the study, and finally, the results.

METHOD

Participants

Participants were 39 English native speaker undergraduates from the United States (33), Canada (4), and Australia (2) studying abroad at three private universities in Buenos Aires, Argentina. Table 1 provides detailed information about the participants.

Data Collection

Data collection occurred in a pretest–posttest format. Participants completed the pretest soon after their arrival in Buenos Aires and the posttest approximately 3 months later (mean number of days between pretest and posttest = 89.5; $SD = 7.4$), also in Buenos Aires. Other than the L1 monologue tasks, all tasks were the same both in the pre- and the posttest.

Tasks and Analyses

Monologue Tasks. The data for the CAF analysis were collected in monologue tasks, 3 in Spanish and 3 in English. Although monologues have the disadvantage of being somewhat artificial, they do avoid the variability that interacting with

TABLE 2
Monologue Tasks

Task Type	Prompt for L2 (Spanish) Task	Prompt for L1 (English) Task
1. Personal activities	Describe what you do on a typical weekday during the school year.	Describe what you do on a typical weekend during the school year.
2. Explain advantages and disadvantages	Explain the advantages and disadvantages of working part time while in college.	Explain the advantages and disadvantages of going to college immediately after graduating from high school.
3. Narrate in past tensee[11]	Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.	Look at the series of pictures below. Using the pictures, tell the story as a sequence of events that occurred in the past.

another person introduces into the task (Segalowitz, 2010). Table 2 presents the details for the monologue tasks.

For each of the monologue tasks, the instructions were presented in SuperLab 5.0. Following the presentation of the instructions, participants were given 30 seconds to prepare their answer and had a maximum of 2 minutes to respond to the prompt. To verify the similarity of the pairs of tasks across languages, each was piloted with three native speakers of English and four native speakers of Spanish. In what follows we present the analyses used for each CAF dimension.

Complexity Analysis. To operationalize the construct of complexity, the present study includes lexical and syntactic measures. To assess lexical variety, participants' full responses to the monologue tasks were transcribed. In cases where responses to a single prompt were more than 100 words in length, transcription was stopped at the end of the T-unit in which they spoke the 100th word (occurred in 25 of 117 pretest responses and 44 of 117 posttest responses). Repetitions and material that was reformulated[5] were not included in the transcriptions. Then, *D* values were obtained using the VocD program in CLAN.

In order to calculate measures of syntactic complexity, participants' transcribed responses were segmented into T-units, and subordinate clauses were marked. Then, the mean number of words per T-unit and the mean number of subordinate clauses per T-unit were calculated. Additionally, composite syntactic complexity scores were calculated by combining the two measures using T scores.

To calculate the measure of lexical complexity, the Guiraud advanced index was used. The Guiraud advanced index is a modified type–token ratio calculated by dividing the number of types outside of the 2,000 most common words in a language by the square root of the number of tokens (Daller, Van Hout, & Treffers–Daller, 2003). The *Corpus de Referencia del Español Actual* (CREA) was used to determine the 2,000 most common words in Spanish. One modification was made: Although the CREA counts singular and plural forms of words separately, if either the singular or plural form of a noun or adjective appeared in the list of the 2,000 most common words, both forms were considered to be among the 2,000 most common words for the purposes of the study.

Accuracy Analysis. To calculate a measure of accuracy, all errors—both lexical and grammatical—were marked in the transcripts, and then the number of errors per 100 words was calculated. The use of lexical items that were imprecise yet logical in the context, such as *bolsa* ('bag') instead of *cesta* ('basket') in a picture-narration task about a picnic, was not considered an error. Additionally, minor mispronunciations, such as *marmelada* instead of *mermelada* ('jam,' 'marmalade'), were not marked as errors. To determine reliability in identifying errors, a second rater scored approximately 20% of the data (15 participants' pretest or posttest responses, totaling 3799 words). The rate of agreement was 96.6% (Kappa = .773).

Fluency Analysis. Thirty-two excerpts were taken from all speaking tasks in both English and Spanish. Typically, each excerpt was taken from second 5 to second 35 of each task. Beginning at second 5 eliminated fillers (*umm, so, okay, well,* etc.) that were sometimes present prior to the start of content that addressed the prompt. If a participant spoke for at least 31 but less than 35 seconds in response to a given prompt, the excerpt was taken from second 1 to second 31. In cases of short responses such as these, fillers were rarely present at the beginning of the sample. In a few cases (8 of 234 pretest responses and 2 of 117 posttest responses), participants spoke for less than 30 seconds in response to a prompt. In these cases, additional time was taken from the preceding or following task so that the excerpts for each participant totaled exactly 90 seconds in each language.

Additionally, following Riggenbach (1991, 2000), unfilled pauses were limited to 3 seconds to prevent any one pause from having undue influence on the results.[6] Prior to taking the 30-second excerpts, pauses longer than 3 seconds were reduced to 3 seconds. There were a total of 15 such pauses in the 234 pretest responses and 7 in the 117 posttest responses.

Once the excerpts were taken, pauses ≥ 0.25 seconds were marked in each excerpt using Praat 5.3.68 (Boersma & Weenink, 2014). As a first step, a script (Lennes, 2002) was used to mark silent pauses. These silent pauses were then verified and adjusted manually as needed, and all filled pauses were marked manually. Each pause was labeled as unfilled (u) or filled (f), long (l; ≥ 0.5 seconds) or short (s; ≥ 0.25 seconds but <0.5 seconds), and mid-clause (m) or end-of-clause (e). Pauses were considered end-of-clause pauses not only if they occurred between clauses as formally defined, but also if they occurred at a natural break, where a comma would be found in written text (see Appendix C for examples).

We also listened to each excerpt and counted the number of syllables and instances of repair. After completing these steps, the following measures of fluency were calculated: (a) *Articulation rate*, that is, the number of syllables per minute, with pauses lasting 0.25 seconds or more removed from the calculation (using the articulation rate, rather than the overall speech rate with pauses included, separates the speed of delivery from pausing [De Jong et al., 2012]); (b) *Repair* per 100 syllables, including repetitions (i.e., number of exact duplications of the same syllable, word, or phrase [Iwashita, Brown, McNamara, & O'Hagan, 2008]), replacements (i.e., number of times one lexical item is substituted for another [Inoue, 2010]), reformulations (i.e., number of grammatical self-corrections [Iwashita et al., 2008]), and false starts (i.e., number of times a speaker abandons an idea without completing it [Inoue, 2010]); (c) *Pauses*, calculated in terms of percent pausing time (i.e., total percent of speaking time taken up by filled or unfilled pauses lasting 0.25 seconds or more), mean length of pause, rate of all pauses (i.e., number of pauses lasting 0.25 seconds or more, per 100 syllables[7]), rate of mid-clause pauses, rate of end-of-clause pauses, rate of long pauses (i.e., number of pauses lasting 0.5 seconds or more, per 100 syllables), rate of short pauses (i.e., number of pauses lasting between 0.25 seconds and 0.49 seconds, per 100 syllables), rate of filled pauses, and rate of unfilled pauses; (d) *Mean length of run* in terms of mean number of syllables spoken between pauses (filled or unfilled) 0.5 seconds or longer; and (e) *Composite fluency scores* (T scores) based on articulation rate, mean length of run, percent of time spent pausing, rate of all pauses, and rate of mid-clause pauses.[8]

Language Knowledge and Language Processing Tasks. As previously discussed, speaking an L2 involves both L2 knowledge and the ability to draw on this knowledge efficiently under time pressure, which requires the gradual automatizing of the subcomponents that together underlie a complex skill such as speaking. The next set of tasks that we present addresses different aspects of linguistic knowledge and processing that underlie CAF.

To measure grammatical knowledge, participants completed a 30-item untimed grammar test (see Appendix A for sample questions). Each item consisted of a sentence containing a grammatical error (verb tense, aspect or mood, adjective–noun agreement, incorrect pronoun), and participants had to detect it and write the correct form of the word in a space below the sentence. The knowledge measured in this task represents the understanding of L2 syntax upon which speakers must draw to produce syntactically accurate and complex speech.

Participants also completed a 30-item untimed vocabulary test, adapted from the Diploma de Español como Lengua Extranjera (DELE). Portions of this test have previously been used in SLA research to assess the proficiency level of nonnative speakers of Spanish (Slabakova, Rothman, & Kempchinsky, 2011; White, Valenzuela, Kozlowskaâ€“MacGregor, & Leung, 2004). This task measures the L2 vocabulary knowledge from which speakers must draw to produce lexically accurate and complex speech.

To measure processing speed, participants completed two timed processing tasks, a picture-naming task and a sentence–picture verification task. These tasks measured participants' ability to efficiently draw on their lexical and syntactic knowledge under time pressure, which is necessary to fluently produce accurate and complex speech. In the picture-naming task, participants were presented a series of 50 pictures of common objects and instructed to name each one in Spanish as quickly as possible. The pictures were presented in SuperLab, and participants' responses were recorded in Audacity. Pictures were selected from the Snodgrass & Vanderwart (1980) picture set, which has been used in a variety of L1 (Bonin, Chalard, Méot, & Fayol, 2002; Hartsuiker & Notebaert, 2010) and L2 (De Jong et al., 2012; De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2013; Sunderman & Kroll, 2009) research. The pictures included were among those found to have at least 90% naming agreement among native speakers of Spanish (Cuetos, Ellis, & Alvarez, 1999; Sanfeliu & Fernandez, 1996). The second processing task was a sentence–picture verification task. Participants heard a sentence and had to indicate whether the sentence accurately described the picture they were looking at. The goal of this task was to measure morphosyntactic processing abilities under time pressure. Sentence–picture verification tasks have been used extensively in L1 research (cortical activation during sentence processing [Mack, Meltzerâ€“Asscher, Barbieri, & Thompson, 2013; Neubauer, Freudenthaler, & Pfurtscheller, 1995] and the effect of aging on sentence comprehension [López–Higes Sánchez et al., 2008]). While sentence–picture verification tasks typically examine vocabulary processing, our goal was to measure speed of morphosyntactic processing, which meant that the task was slightly modified from that used in previous studies. For example, participants heard the sentence *Mira la televisión* ("He/she watches television"). A mismatched picture showed two people watching television. A matching picture would show one person watching television (see Appendix B for more examples).

In the present study, the sentences were presented in auditory form since our intent was to represent online speaking processing as realistically as possible. Participants heard a sentence, immediately followed by the presentation of a picture. Reaction time was measured from the point at which the picture appeared. The task included 34 sentence–picture pairs in the main set and 6 pairs in the warm-up set.

Each item in both the picture-naming and grammar-processing tasks was analyzed using the pretest data. In the picture-naming task, 7 of 50 items had an accuracy rate of 75% or less. These items were removed for all participants, and the mean accuracy rate for the remaining 43 items was 94.80% on the pretest. In the sentence–picture verification task, 6 of 34 items had an accuracy rate of 70% or less. These items were removed for all participants, and the mean accuracy rate for the remaining 28 items was 86.24% on the pretest.

For the picture-naming task, certain data points were also eliminated for individual participants. Following Sunderman & Kroll (2009) and De Jong et al. (2012, 2013), outliers (reaction times greater than two standard deviations above or below each participant's mean) were excluded. As Sunderman and Kroll (2009) explain, very short reaction times suggest the influence of "anticipatory processes" (p. 86), whereas very long times may reflect lapses of attention. This process led to the removal of a total of 74 data points (4.30% of the data) on the pretest and 74 data points (4.41% of the data) on the posttest. Additionally, 0.99% of the pretest data and 1.49% of the posttest data had to be eliminated due to false activation of the voice key or failure of the voice key to detect a response.

After these adjustments were made, the mean reaction time in milliseconds was calculated for each participant's correct responses. All participants approached or exceeded 75% accuracy on the pretest. One participant was eliminated from the pretest and two from the posttest due to mistakes following instructions.

For the sentence–picture verification task, participants' reaction times were log transformed prior to calculating a mean score for correct responses for each participant. This transformation was performed to correct for positively skewed data, following previous researchers using reaction time measures (Bott & Noveck, 2004; Edwards & Lahey, 1996). On both the pretest and posttest, three participants responded with less than 75% accuracy, and their data for this task were not used.

All Data. Prior to carrying out correlational analyses, data were tested for normality using the Shapiro–Wilk test. Most data were found to be normally distributed; log transformations were applied to non-normally distributed data, including pretest picture-naming scores and change scores for grammar and vocabulary.

Research Questions

The study focused on the following research questions:

RQ1. Are there changes in CAF, linguistic knowledge, and/or linguistic processing speed from pre- to post study abroad?

RQ2. Is there a relationship between participants' L2 fluency, accuracy, and complexity measures, either pre-SA or in the amount of change during SA?

RQ3. Does participants' linguistic knowledge and/or linguistic processing speed predict their L2 complexity, accuracy, and fluency, either pre-SA or in the amount of change during SA?

Research Question 1 considers what, if any, changes occurred in participants' speaking ability, linguistic knowledge, and linguistic processing speed over the course of the 3-month SA session. Research Question 2, on the other hand, examines how the CAF dimensions relate to each other pre-study abroad and post-study abroad. Finally, Research Question 3 considers how the linguistic processing and knowledge measures relate to CAF. In the following section, we present the results addressing each of these questions in turn.

RESULTS

RQ1: Changes in CAF, Linguistic Knowledge, and Processing Speed

Table 3 summarizes mean pretest and posttest scores on each complexity, accuracy, and fluency measure. Paired t-tests were used to check for significant differences between pretest and posttest scores.

As Table 3 shows, there were significant changes in participants' fluency from the pretest to the posttest on a number of measures.[9] However, the rate of end-of-clause pauses and the rate of short pauses did not significantly change. Looking at rate of repair, there was no significant change.

Regarding accuracy, the mean number of errors per 100 words significantly decreased from pre- to post-SA. Participants also experienced small[10] but significant gains on both measures of syntactic complexity, the mean number of subordinate clauses per T-unit, and the mean number of words per T-unit. Additionally, participants experienced significant gains in lexical complexity; however, gains in lexical variety were not significant.

As Table 4 shows, participants also experienced significant gains on the linguistic knowledge measures of grammar and vocabulary knowledge as well as the tasks measuring L2 processing speed. They experienced large gains on the picture-naming task ($d = .98$), moderate gains on the grammar ($d = .53$) and vocabulary ($d = .68$) tests, and small gains on the sentence–picture verification task ($d = .22$).

RQ2: The Relationship Between Fluency, Accuracy, and Complexity Pre- and Post-SA

Pearson correlations were used to examine the relationship between fluency, accuracy, syntactic complexity, and lexical complexity in the pretest data. As Table 5 shows, there were moderate correlations between pre-SA fluency and accuracy, fluency, and lexical complexity, and accuracy and lexical complexity. Correlations with syntactic complexity were not significant. Correlations involving fluency were repeated with a control for L1 fluency, given the possibility that differences in L2 fluency may be partly due to differences in individuals' general speaking style (Segalowitz, 2010). However, controlling for L1 fluency did not significantly change the correlations.

Correlations were also used to examine the relationship between the amount of change in fluency, accuracy, syntactic complexity, and lexical complexity during the semester abroad. As Table 6 shows, none of these correlations were significant.

RQ3: Linguistic Knowledge and Processing Speed as Predictors of CAF

Stepwise regression was used to examine whether pre-SA grammar and vocabulary scores and measures of processing speed predicted pre-SA fluency, accuracy, lexical variety, and complexity. As Table 7 indicates, participants' scores on the written grammar measure were a significant predictor of pre-SA accuracy, lexical complexity, and lexical variety. Sentence–picture verification and picture-naming, both measures of processing speed, were significant predictors of pre-SA fluency. Picture-naming was also a predictor of lexical complexity.

Stepwise regression was also used to examine whether pre-SA grammar and vocabulary scores and measures of processing speed predicted changes in fluency, accuracy, lexical variety, and complexity during study abroad. To account for the fact that there was an inverse relationship between participants' pre-SA fluency, accuracy, lexical variety, and complexity and the amount of change during SA, their pre-SA scores on these measures were entered as independent variables as well. As Table 8 shows, the strongest predictor of change on any given measure during study abroad was participants' pre-SA scores on that measure; for example, participants who had lower fluency scores pre-SA tended to experience greater gains in fluency during SA. However, pre-SA vocabulary and picture-naming

TABLE 3
Mean Pretest and Posttest Scores for all Speaking Measures

	Mean Pretest Score	*SD*	Mean Posttest Score	*SD*	Sig. (2-tailed)	Effect Size (*d*)
Fluency						
Syllables per second	3.90	.68	4.08	0.58	.002	.28
Percent pausing time	39.59	11.16	35.47	10.01	<.001	.39
Rate[10] of all pauses	16.84	5.85	14.83	4.64	<.001	.38
Rate of mid-clause pauses	8.94	4.61	7.03	3.55	<.001	.46
Rate of end-of-clause pauses	7.90	2.08	7.80	2.11	.780	.05
Rate of long pauses (≥ .50 seconds)	12.98	5.78	10.85	4.38	<.001	.42
Rate of short pauses (.25–.49 seconds)	3.87	1.58	3.99	1.63	.691	.07
Rate of unfilled pauses[11]	18.78	7.75	15.79	5.78	<.001	.44
Rate of filled pauses	6.56	4.31	4.33	3.60	<.001	.56
Rate of repair	3.48	2.34	3.43	2.75	.863	.02
Mean length of run between pauses ≥ 0.50 seconds	8.81	3.74	10.55	5.02	.005	.39
Accuracy						
Errors per 100 words	10.02	5.63	7.01	4.09	<.001	.61
Syntactic complexity						
Subordination (subordinate clauses per T-unit)	.44	.17	.52	.23	.028	.40
Mean length of T-unit (in words)	11.58	2.13	12.64	2.80	.029	.43
Lexical complexity						
Guiraud advanced index	1.40	.54	1.69	.51	<.001	.55
Lexical variety						
VocD	66.83	14.41	69.75	13.42	.107	.21

TABLE 4
Mean Pretest and Posttest Scores for Grammar, Vocabulary, and Processing Measures

	Mean Pretest Score	*SD*	Mean Posttest Score	*SD*	Sig. (2-tailed)	Effect Size (*d*)
Grammar (out of 30 points)	18.71	4.64	21.00	3.99	<.001	.53
Vocabulary (out of 30 points)	20.63	5.00	23.76	4.19	<.001	.68
Sentence–picture verification						
Log values	3.32	.14	3.29	.15	.039	.22
Converted ms	2107		1957			
Picture-naming (ms)	1327	264	1112	160	<.001	.98

scores were also significant predictors. Higher pre-SA vocabulary scores predicted greater gains in accuracy, syntactic complexity, and lexical complexity, while higher pre-SA picture-naming scores predicted greater gains in lexical variety.

Additionally, stepwise regression was used to examine whether changes in grammar and vocabulary scores and measures of processing speed during SA predicted changes in fluency, accuracy, lexical variety, and complexity. As Table 9 indicates, changes in picture-naming scores predicted greater gains in accuracy; changes in fluency, complexity, and lexical variety could not be predicted from changes in grammar, vocabulary, or processing speed.

DISCUSSION

The goal of the present study was to investigate how speaking proficiency develops in a group

TABLE 5
Correlations Among Fluency, Accuracy, and Lexical Variety, and Complexity Pretest Scores

	Accuracy	Syntactic Complexity	Lexical Complexity	Lexical Variety
Fluency	.610**	.173	.528**	.609**
With control for L1 fluency	.620**	.124	.555**	.556**
Accuracy		.186	.619**	.427**
Syntactic complexity			−.007	−.085
Lexical complexity				.700**

Note. *$p < .05$; ** $< .01$

TABLE 6
Correlations Among Fluency, Accuracy, Lexical Variety, and Complexity Change Scores

	Accuracy	Syntactic Complexity	Lexical Complexity	Lexical Variety
Fluency	.174	.082	−.002	.151
Accuracy		.140	.122	−.245
Syntactic complexity			.168	.173
Lexical complexity				.157

Note. *$p < .05$; **$p < .01$

TABLE 7
Predictors of Pre-SA Fluency, Accuracy, Complexity, and Lexical Variety

Dependent Variable	Significant Predictors	*B* (beta)	*t*	Total R^2	Total *F*	*p*
Fluency	Sentence–picture verification	−.533	−3.859	.500	15.97	<.001
	Picture-naming	−.290	−2.097			
Accuracy	Grammar	−.720	−5.963	.519	35.56	<.001
Syntactic complexity	No predictors					
Lexical complexity	Grammar	.514	3.622	.588	22.81	<.001
	Picture-naming	−.338	−2.382			
Lexical variety	Grammar	.653	4.955	.427	24.56	<.001

of native English speakers participating in a 3-month study abroad program in Buenos Aires, Argentina. Specifically, we examined how dimensions of CAF relate to each other, how linguistic knowledge and processing relate to CAF, and how all these subcomponents of speaking proficiency changed over the course of the SA program.

Changes in CAF, Linguistic Knowledge, and Processing Speed

In our first research question, we asked whether there were changes in participants' L2 fluency, accuracy, syntactic and lexical complexity, lexical variety, grammar and vocabulary knowledge, and processing speed during SA. Overall, our findings show that the response to this question is affirmative. The present study's finding of an increase in syntactic complexity is in line with the findings of two previous studies (Lennon, 1990b; Serrano et al., 2012), as is the finding of no significant change in lexical variety (Mora & Valls Ferrer, 2012; Serrano et al., 2012). Unlike previous studies, the present study also included a measure of lexical complexity, finding a significant increase from pre- to post-SA. Particularly for more advanced learners, it may be that SA provides an opportunity to incorporate a greater number of low-frequency words into their active vocabulary (thus increasing their lexical complexity), but this change does not necessarily lead to a significant increase in their overall lexical variety.

The finding that participants experienced significant gains in accuracy is in line with the results of other recent studies (Llanes & Muñoz, 2009; Mora & Valls Ferrer, 2012; Serrano et al., 2012). Together, these results challenge the

TABLE 8
Pre-SA Predictors of Change in Fluency, Accuracy, Complexity, and Lexical Variety

Dependent Variable	Significant Predictors	*B* (beta)	*t*	Total R^2	Total *F*	*p*
Fluency	Pre-SA fluency	−.554	−3.828	.307	14.65	.001
Accuracy	Pre-SA accuracy	−.893	−5.530	.498	15.86	<.001
	Pre-SA vocabulary	.430	2.663			
Syntactic complexity	Pre-SA syntactic complexity	−.461	−3.385	.411	11.15	<.001
	Pre-SA vocabulary	.409	3.005			
Lexical complexity	Pre-SA lexical complexity	−.869	−5.258	.465	13.88	<.001
	Pre-SA vocabulary	.587	3.550			
Lexical variety	Pre-SA lexical variety	−.641	−3.501	.277	6.13	.006
	Pre-SA picture-naming	.379	2.067			

TABLE 9
During-SA Predictors of Change in Fluency, Accuracy, Complexity, and Lexical Variety

Dependent Variable	Significant Predictors	*B* (beta)	*t*	Total R^2	Total *F*	*p*
Fluency	No predictors					
Accuracy	Change in picture-naming	.371	2.150	.137	4.62	.04
Syntactic complexity	No predictors					
Lexical complexity	No predictors					
Lexical variety	No predictors					

assertion that, although SA may lead to large gains in fluency, it does not have much impact on accuracy (DeKeyser, 2010; Freed, 1998). In the present study, the effect size for gains in accuracy ($d = .61$) was actually slightly greater than the effect sizes for gains in fluency ($d = .28$ to .56).

Regarding repair (one measure that has led to discrepancies across previous studies), the present study found no significant changes, but if participants are divided into two groups according to their pre-SA fluency scores, there is a trend toward higher rates of repair for those who began with fluency scores below the mean (4.23 repairs per 100 syllables pre-SA; 4.51 post-SA) and a trend toward lower rates of repair (2.87 repairs per 100 syllables pre-SA; 2.35 post-SA) for those who began with fluency scores above the mean. Mora and Valls Ferrer (2012), who found a decrease in repair, had participants who were advanced L2 learners in a translation and interpreting program, whereas D'Amico (2010), who found an increase in repair, had participants who were enrolled in an intermediate-level course in the L2. Looking at the data from these two studies together with the present study suggests, as an initial hypothesis, that frequency of repair may develop in a non-linear manner and changes may therefore differ for participants at different stages of their L2 learning.

The present study also included additional measures of fluency not commonly considered in other SA research. For example, we included separate measures for mid-clause and end-of-clause pauses. The finding that the rate of mid-clause pauses significantly decreased from pre- to post-SA but the rate of end-of-clause pauses did not decrease suggests that the overall reduction in pausing occurred for pauses that are less natural, given that pausing at the end of clauses is more common in L1 speech in general (Chambers, 1997; Davies, 2003), a fact confirmed by the L1 data collected from the participants in this study as well. Table 10 provides the L1 and L2 pause data:

As Table 10 shows, a notable difference between participants' L1 and L2 fluency prior to SA was that in the L1 they were often able to formulate and articulate full clauses without stopping mid-clause), whereas in their L2, they paused more frequently in the middle of clauses than at the end of clauses (suggesting that they had difficulty formulating and articulating full clauses without stopping mid-clause). Post-SA, their pattern of mid-clause and end-of-clause pauses in the L2 approximated the L1 pattern, indicating important changes from pre- to post-SA.

In summary, our results confirm the findings of previous studies that L2 learners experience gains in fluency, accuracy, and syntactic complexity

TABLE 10
Mid-Clause and End-of-Clause Pauses, L1 and L2

	Mean (*SD*) L1 Score	Mean (*SD*) L2 Pretest Score	Mean (*SD*) L2 Posttest Score
Rate of mid-clause pauses	3.57 (1.98)	8.94 (4.61)	7.03 (3.55)
Rate of end-of-clause pauses	7.53 (1.52)	7.90 (2.08)	7.80 (2.11)

during SA. We also took a more detailed look at changes in fluency and found a reduction in pausing specifically for mid-clause pauses, indicating a post-SA pattern of pausing in the L2 that more closely approximated that of the L1. Additionally, by analyzing not only lexical variety but also lexical complexity, we found evidence of significant changes in lexis during SA.

The Relationship Between Fluency, Accuracy, and Complexity

We also examined the relationship between fluency, accuracy, and syntactic and lexical complexity, both pre-SA and in terms of changes during SA. The finding of significant correlations between fluency, accuracy, and lexical complexity in the pre-SA data suggests that in the long term, these dimensions of speaking ability all develop together to some degree. That is, a learner who has reached a certain level of fluency is likely to have also reached a certain level (within a range; not a precise level) of accuracy and lexical complexity. Of course, this finding does not address the question of whether and how they are interdependent or whether they all simply tend to increase over time with more exposure to the L2.

The finding that there were no significant correlations between changes in fluency, accuracy, and syntactic and lexical complexity during SA suggests that, in the short term, the development of each of these dimensions of speaking ability is not related to the development of the other dimensions. For the group of participants as a whole, there is no evidence of either trade-off effects between dimensions (i.e., greater development in one dimension associated with less development in another dimension) or positive relationships between dimensions (i.e., greater development in one dimension associated with greater development in another dimension).

Linguistic Knowledge and Processing Speed as Predictors of CAF

In our third research question, we asked whether participants' knowledge of L2 grammar and vocabulary and/or L2 processing speed predicted their L2 fluency, accuracy, lexical variety, syntactic complexity, and/or lexical complexity, both pre-SA and in terms of changes during SA. The finding that higher pre-SA vocabulary and picture-naming scores predicted greater gains in accuracy, lexical variety, and syntactic and lexical complexity suggests that having more developed linguistic knowledge and language processing abilities prior to SA gives learners a slight advantage in accuracy, lexical variety, and complexity gains during SA, but not in fluency gains.

One explanation for this finding is that, as Lennon (1990a) suggests, fluency may be fundamentally different from other dimensions of speaking in that accuracy and complexity are directly related to linguistic knowledge, but fluency is not. It would appear that fluent speech in an L2 requires at least a basic knowledge of L2 vocabulary, for one could hardly be said to speak fluently in a given language without knowing the language at all. However, if a learner is relatively unconcerned about errors, he or she could potentially speak quite fluently (with *fluency* defined according to temporal measures) despite limited linguistic knowledge. The same is not true of accuracy, complexity, and lexical variety. The present study's finding that the strongest predictor of pre-SA fluency was a measure of processing speed (sentence–picture verification), whereas the strongest predictor of accuracy and complexity was a measure of linguistic knowledge (a written grammar test), supports this view that fluency depends—at least to some degree—on different underlying skills than do other dimensions of speaking.

In a SA setting, one of the presumed benefits is the chance to speak the L2 more frequently than in an at-home classroom setting by using the language in daily life. Given this opportunity to practice speaking, it may be that participants were able to experience gains in fluency regardless of their pre-SA linguistic knowledge, since gains in fluency reflect a speeding up of speaking performance—which could also be described as an increase in the automaticity that

~~Level~~ (1989) argues is necessary for uninterrupted speech—and not necessarily any other changes in that performance. The formulation stage may be automatized without necessarily requiring a high degree of accuracy or complexity (and therefore not necessarily requiring a high level of linguistic knowledge). However, participants with higher levels of linguistic knowledge were also able to experience greater gains in accuracy and complexity during their time abroad by increasingly putting that knowledge into practice, as reflected in changes in word choice, verb conjugations, and sentence structure in their speaking performance.

Based on these findings, one might also expect that changes in linguistic knowledge and processing speed during SA would predict changes in fluency, accuracy, complexity, and lexical variety during SA. However, we found only one significant relationship in this regard; changes in picture-naming scores predicted greater gains in accuracy, but changes in fluency and complexity could not be predicted from changes in linguistic knowledge and processing speed. We consider two possible explanations for these results.

First, in the case of changes in grammar and vocabulary, the lack of a significant relationship to changes in fluency, accuracy, and complexity may be partly a measurement issue. Since the maximum score on each of these measures was 30 points, the range of change was fairly limited: −4 to +9 on the grammar measure and −1 to +9 on the vocabulary measure.

Second, the duration of the study (3 months) may have been too short a period of time to observe a relationship between changes in participants' linguistic knowledge and changes in their speaking abilities. Language teachers can testify to the fact that gaining explicit knowledge of a language feature and being able to produce it correctly without time pressure does not mean that learners' spontaneous speech immediately reflects that new knowledge. As previously mentioned, speaking requires not only declarative knowledge, but also the ability to process knowledge quickly (De Jong et al., 2012), which develops over time with practice. It is worth noting that the one significant predictor of gains in accuracy was an increase in processing speed (faster picture-naming times), not the gaining of new linguistic knowledge (increases in grammar and vocabulary scores). It seems, then, that learners who have gained greater linguistic knowledge over time *prior* to SA are slightly better positioned to experience gains in speaking ability during SA as compared to their peers, but the effect may not be the same for gains in linguistic knowledge *during* SA. In practical terms, this finding suggests that learners who plan to spend a relatively short period of time abroad, such as the one semester sojourn in this study, and hope to experience significant gains in speaking ability during their SA should be encouraged to make a strong effort to learn L2 grammar and vocabulary prior to SA (and not to assume that most real L2 learning takes place abroad, as some students tend to do). Future research with participants studying abroad for a full academic year would be helpful in determining whether or not the same applies to longer-term SA.

CONCLUSIONS

In the present study we examined how multiple dimensions of CAF change over the course of a 3-month SA session and crucially, whether these changes are related to linguistic knowledge and processing speed, two factors that underlie the formulation stage of Levelt's speech production model. More specifically, we hypothesized that changes in CAF would be positively related to the ability of L2 speakers to process input (lexical retrieval and grammatical processing) in real time and also the stored linguistic knowledge that they can draw upon when time is not an issue. Our findings revealed that, if learners' pre-SA levels of L2 speaking ability are taken into account, participants who began their SA experience with higher levels of linguistic knowledge and processing abilities tended to experience greater gains in accuracy and syntactic and lexical complexity over the 3-month study abroad session. These results suggest that having greater vocabulary and faster processing skills may 'free up' resources that allow learners to improve their accuracy and complexity in spoken language. Interestingly, fluency did not show this relationship.

Acknowledgments

This research was supported by grants from Language Learning's Small Grants Research Program, the Executive Council of Graduate and Professional Students at the University of Iowa, and the Foreign Language Acquisition Research and Education program at the University of Iowa. The authors would like to thank the faculty and staff associated with the International Office at the Pontificia Universidad Católica Argentina in Buenos Aires for their assistance with this project.

NOTES

[1] Analyzing these relationships in a study abroad context allows us to control for factors such as motivation, since all participants elected to study abroad in Buenos Aires, and it also maximizes the possibility that changes will be observed and thereby permit a clearer examination of the relationships considered here.

[2] An AS-unit is "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster, Tonkyn, & Wigglesworth, 2000, p. 365).

[3] A T-unit is "one main clause plus whatever subordinate clauses happen to be attached to or embedded within it" (Hunt, 1965, p. 305).

[4] Following work by Grey, Cox, Serafini, & Sanz, (2015), we used participants as their own control and employed a within-subjects design. Using an at-home control group runs the risk of introducing variables such as motivation and self-selection as well as different amounts of language contact outside of class. Moreover, we were interested in how fluency development correlates with other aspects of language development such as processing and lexical/grammatical knowledge, not in how speaking changes in a study abroad context versus a stay-at-home context, per se.

[5] For example, if a participant initially produced the wrong form of a verb and then corrected it, the wrong form initially produced was not included in the transcription.

[6] No filled pauses exceeded 3 seconds.

[7] It could be argued that measuring pauses per 100 syllables rather than pauses per minute conflates pausing and speech rate; however, the two cannot be completely separated in calculating a rate of pauses. Measuring pauses per minute does not take into account the fact that a participant who speaks many syllables per minute could potentially pause many more times per minute than a participant who speaks few syllables per minute (since there must be at least one syllable spoken between pauses), especially when measuring pauses as short as .25 seconds.

[8] Since it seemed that including all eight measures of pausing in the composite score would give too much weight to pausing in subsequent quantitative analyses, percent of time spent pausing and rate of all pauses were chosen as the most general measures of pausing. Rate of mid-clause pauses was also included, as it was possible for participants to have a significant shift in the location of their pauses from pre- to post-SA without this change necessarily being reflected in either of the more general measures of pausing.

[9] Participants narrated stories based on pictures from Heaton (1966).

[10] Rates are per 100 syllables.

[11] The rate of unfilled pauses and rate of filled pauses add up to more than the rate of all pauses because of the way the pauses were counted. For the rate of all pauses, mid-clause pauses, end-of-clause pauses, long pauses, and short pauses, instances of a filled pause immediately followed by an unfilled pause (or vice versa) were counted as a single pause. For the rate of unfilled pauses and filled pauses, filled and unfilled pauses were all counted as separate pauses.

[12] It should be noted that the use of multiple t-tests increases the likelihood of a Type I error. However, given that the results for more than half of the measures of fluency are significant at $p < .001$, it seems safe to conclude that the data demonstrate a real change in fluency.

[13] Effect sizes are interpreted following Cohen (1988).

REFERENCES

Boersma, P., & Weenink, D. (2014). Praat 5.3.68 [Computer software]. Accessed 25 March 2014 from http://www.praat.org

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*, 89–114.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*, 535–544.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, *31*, 650–658.

D'Amico, M. L. (2010). *The effects of intensive study abroad and at home language programs on second language acquisition of Spanish*. (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

Daller, H., Van Hout, R., & Treffersâ€“Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*, 197–222.

Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon, UK: Multilingual Matters.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*, 5–34.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*, 893–916.

DeKeyser, R. (2007). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208–226). New York: Cambridge University Press.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. *Foreign Language Annals*, *43*, 80–92.

DeKeyser, R. (2016). Of moving targets and chameleons. *Studies in Second Language Acquisition*, *38*, 353–363.

Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech and Hearing Research*, *39*, 1263–1273.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375.

Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *4*, 31–60.

Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *Modern Language Journal*, *99*, 137–157.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*, 169–177.

Heaton, J. B. (1966). *Composition through pictures*. London: Longman.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Philadelphia/Amsterdam: John Benjamins.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 461–473.

Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal*, *54*, 300; 305–309.

Inoue, C. (2010). Investigating the sensitivity of the measures of fluency, accuracy, complexity, and idea units with a narrative task. In G. Bota, H. Hargreaves, L. Chia–Chun, & R. Rong (Eds.), *Papers from the Lancaster University postgraduate conference in linguistics & language teaching* (pp. 129–153). Lancaster, PA: Lancaster University.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*, 24–49.

Lennes, M. (2002). Mark pauses [Praat script]. Accessed 25 March 2014 at http://www.helsinki.fi/~lennes/teaching/aisv/stuff/mark_pauses.praat

Lennon, P. (1990a). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *3*, 387–417.

Lennon, P. (1990b). The advanced learner at large in the L2 community: Developments in spoken performance. *International Review of Applied Linguistics in Language Teaching*, *28*, 309–324.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Llanes, Á., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*, 353–365.

López-Higes Sánchez, R., Rubio Valdehita, S., Martín Aragoneses, M. T., & Del Río Grande, D. (2008). Variabilidad en la comprensión gramatical en el envejecimiento normal. [Variability in grammatical comprehension in normal aging.] *Revista de Logopedia, Foniatría y Audiología*, *28*, 15–27.

Mack, J. E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain Sciences*, *3*, 1198–1214.

Mora, J. C., & Valls Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*, 610–641.

Neubauer, A., Freudenthaler, H. H., & Pfurtscheller, G. (1995). Intelligence and spatiotemporal patterns of event-related desynchronization (ERD). *Intelligence*, *20*, 249–266.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*, 557–581.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*, 423–441.

Riggenbach, H. (2000). *Perspectives on fluency*. Ann Arbor, MI: The University of Michigan Press.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, *28*, 537–555.

Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA Project, an appraisal. In C. Pérez–Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 1–17). Philadelphia/Amsterdam: John Benjamins.

Sanz, C., & Grey, S. (2015). Effects of conditions on L2 development: Moving beyond accuracy. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 301–324). Philadelphia/Amsterdam: John Benjamins.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge/Taylor & Francis\.

Serrano, R., Tragant, E., & Llanes, Á. (2012). A longitudinal analysis of the effects of one year abroad. *The Canadian Modern Language Review*, *68*, 138–163.

Slabakova, R., Rothman, J., & Kempchinsky, P. (2011). Gradient competence at the syntax–discourse interface. *EUROSLA Yearbook*, *11*, 218–243.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31*, 532–553.

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30*, 79–99.

SuperLab 5.0 [Computer software]. (2014). San Pedro, CA: Cedrus.

Q3 Swender, E., Conrad, D. J., & Vicars, R. (2012). *ACTFL proficiency guidelines*. Alexandria, VA: ACTFL.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics, 40*, 117–150.

White, L., Valenzuela, E., Kozlowskaâ€“MacGregor, M., & Leung, Y. K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics, 25*, 105–133.

APPENDIX A

Sample Questions from Grammar Test

Each sentence below contains a grammatical error. Your task is to find the error and write the corrected form in the box below the sentence.

Model: Adriana están feliz.

Está

3. Se le mandé a las diez de la mañana por correo electrónico.

6. Se prohíben fumar aquí.

13. El médico quien me trató ya no trabaja en el mismo hospital.

APPENDIX B

Samples from Sentence–Picture Verification Task

	Picture	Sentence
Matching pictures and sentences	Nora and Elisa are reading a book.	Leen un libro. *They read a book.*
	Juana is serving drinks to Ana and Manuel.	Juana les sirve bebidas. *Juana serves them drinks.*
Mismatched pictures and sentences	Nina is hugging Micah.	Nina los abraza. *Nina hugs them.*
	Adela is putting Rosa to bed.	Adela se acuesta. *Adela goes to bed.*

APPENDIX C

Examples of Mid-Clause and End-of-Clause Pauses

Mid-clause pauses, English

I'll have [pause] one or two classes during the day.

There was a boy biking on [pause] a narrow road.

Mid-clause pauses, Spanish

La mamá [pause] les dio [pause] un mapa. = *The mother [pause] gave them [pause] a map.*

Yo [pause] como [pause] café con leche. = *I [pause] eat [pause] coffee with milk.*

End-of-clause pauses, English

Taking time off helps them prepare for that [pause] but [pause] I think the advantage is …

If I have time for breakfast [pause] I'll usually make toast.

End-of-clause pauses, Spanish

Mientras estaban caminando hacia el parque [pause] el perro comió todo el picnic. = *While they were walking toward the park [pause] the dog ate the whole picnic.*

Paso un tiempo en la computadora [pause] y [pause] veo una película. = *I spend time on the computer [pause] and [pause] I watch a movie.*

Queries

Q1: Author: Please confirm that given names (red) and surnames/family names (green) have been identified correctly.

Q2: Author: Please check all the footnote as per original manuscript. As while converting note into footnote, numbering has changed.

Q3: Author: The reference SuperLab 5.0 [Computer software] 2014 has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.