WRITE BEFORE YOU SPEAK:
THE IMPACT OF WRITING ON L2 ORAL NARRATIVES

By

Alyssa Bulow

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Teaching English to Speakers of Other Languages — Master of Arts

2020

ABSTRACT

WRITE BEFORE YOU SPEAK:
THE IMPACT OF WRITING ON L2 ORAL NARRATIVES

By

Alyssa Bulow

Current literature suggests that writing may better facilitate language learning than speaking practice alone, but direct empirical research demonstrating this is limited. Evidence is also limited as to whether grammar and vocabulary learned while writing can transfer to speaking. This study investigates the prediction that written planning, even more so than oral planning, leads to improved oral narratives. Thirty-four Spanish-speaking learners of English were randomly assigned to one of two groups: writing rehearsal or oral rehearsal; rehearsal being individual practice before the final task. The writing group composed a story ending in the written modality while the oral group rehearsed by narrating theirs out loud. Both groups recorded their oral story continuation task as the final product. In order to compare the impact of writing versus oral rehearsal on learners' subsequent oral performance, final narratives were examined using complexity, accuracy, and fluency measures. Results showed that the writing group produced more fluent and lexically diverse narratives than the speaking group but there was no effect on accuracy, and limited effects on grammatical complexity. The study concludes with pedagogical implications for using writing tasks to prepare students for oral tasks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

It has been suggested that writing, compared to speaking, can lead to the short-term use and acquisition of more complex and accurate structures by encouraging greater precision, deeper cognitive processes, and greater access to explicit knowledge. This is partially due to the slower pace and less ephemeral nature of writing (Williams, 2012). However, the relationship between the support that writing tasks and practice lend to speaking tasks has not been heavily researched and yet remains unclear (Polio, in press). Furthermore, although there are many studies on the effects of planning on oral and written texts, there is little research on the modality of that planning, specifically of the impact of writing on oral performance tasks. The present study examines how L2 oral production, elicited using a story continuation task (SCT), is affected differently by written and oral practice to test the hypothesis that writing practice yields more linguistically sophisticated, accurate, and possibly more fluent oral narratives.

LITERATURE REVIEW

Because the focus of this study is on the effects of the modality of planning or rehearsal, I will begin with a discussion of the general research on the effects of planning on oral language. While there are also many studies on the effects of planning on written language (for a review see Johnson, 2017), those will not be addressed here. Next, I examine why written planning or rehearsal might better facilitate various aspects of oral production compared to oral rehearsal. I end with a discussion related to the data elicitation technique, namely the story continuation task, and the outcome measures used to assess learning.

The Effects of Pre-Task Planning on Oral Language Production

Information Processing Theory suggests that humans have a limited processing capacity and so cannot attend to all aspects of a task simultaneously. Planning offers a way to enable learners to focus their limited processing capacity. Thus, pre-task (planning *before* completing the task) and online planning (planning occurring in the moment *while* completing the task) have the capability to improve performance. (Ellis 2005b; 2009)

Pre-task planning can be separated further into rehearsal (planning where learners have an opportunity to *perform the complete task* once before performing it again) or strategic planning (planning overall content and lexical items, but *no chance to completely rehearse* the final task) (Ellis 2005b, 2009). To avoid confusion this study uses the term *rehearsal* to refer to the practice taking place during the preparation time for the final task, for both participants in the writing rehearsal group (WR) and the oral rehearsal group (OR). Although the participants could have been using strategic planning or engaging in pre-task rehearsal, the use of this term is intended to encompass both.

Rehearsal, and the opportunity for learners to focus their processing capacity, can be useful for L2 learners as they may find it challenging to attend simultaneously to meaning and form and decide how to allocate their attention by prioritizing certain aspects of production (Yuan & Ellis, 2003; Anderson, 1995; Skehan, 1996; VanPatten, 1990). In general, research lends support to the idea that pre-task planning has a positive impact on language production, particularly regarding fluency and complexity. (Ortega, 1999; Rahimpour & Safarie, 2011; Yuan & Ellis, 2003).

A recent meta-analysis on pre-task planning examines the role of planning on oral tasks and selected 40 studies from 1995-2016 (Suzuki, 2017). Suzuki found that several studies supported the effectiveness of planning on the fluency of L2 learners' oral production (Yuan & Ellis, 2003; Foster & Skehan, 1996; Gilabert, 2007; Ortega, 1999, Sasayama & Izumi, 2012) while the effects of planning on accuracy (Mehnert, 1998; Foster & Skehan, 1999; Lee & Oh, 2007; Mochizuki & Ortega, 2008) and complexity (Bei, 2010; Kawauchi, 2005; Nitta, 2007; Wigglesworth, 1997; Wang & Song, 2015; Yuan, 2001) have yielded mixed results. It has been suggested that this could be partially due to different units being used to measure complexity, accuracy, and fluency (CAF).

Similar to planning, task repetition offers benefits to language learners by creating chances to monitor output in order to possibly increase accuracy or complexity. Ahmadian and Tavakoli (2010) found that the chance to partake in both online planning and task repetition can enhance accuracy, complexity, and fluency to a large extent. Mehnert (1998) looked at different lengths of planning time and found that fluency increases with amount of planning time, but that greater complexity was shown only in the planning group with the most time allotted (10 minutes).

Yuan and Ellis (2003) compared the impact of pre-task, or strategic, planning with online (i.e., moment-by-moment) planning on task performance. In discussing the findings of Ellis'

(1987) study on narrative tasks, they noted that one aspect that sets apart the written from oral modality is that writing allows more opportunity for online planning and is less taxing on working memory than speaking. The participants in Yuan and Ellis were shown a set of pictures and either given an hour to: write a story as their only task; retell their written story orally without access to it (but the ability to record it twice); or to immediately tell the story orally. According to Yuan and Ellis (2003), there is sufficient evidence showing that pre-task planning helps learners produce more fluent and complex language when performing the task; however, whether greater accuracy is promoted remained unclear.

When learners must plan very quickly or are in pressured speech situations, they mainly search for lexical material (Ochs, 1979) rather than grammatical information. Pre-task planners typically start with conceptualization and move on to the formulation stage if time allows (Levelt, 1989). It is suggested that pre-task speech planners will recall what they want to say rather than how to say it (Ellis & Yuan, 2003) thus enhancing complexity and fluency rather than accuracy. However, not all researchers agree on this point, as others have hypothesized that writing, such as the WR group engages in, does encourage attention to linguistic form and may allow for skill transfer from writing to speaking. (Blake, 2009; Payne, 2002; Weissberg, 2000; Williams, 2008).

Johnson, Mercado, and Acevedo (2012) suggest that the Limited Attentional Capacity Model (which assumes attention is limited in capacity and may hamper our ability to carry out simultaneous tasks) and the Cognition Hypothesis (that increasing task complexity influences the quality of L2 production: Jackson & Suethanapornkul, 2013) may not be applicable to writing; if this is indeed the case participants in the current study's writing rehearsal group (WR) could benefit by being less influenced by the task complexity as a result of the written modality.

However, Johnson et al (2012) focused on writing with no bridge to oral language, such as is provided by the story continuation task in the current study. The writing provided in the WR group may facilitate oral production better than mere speaking practice due to planning time free from concerns posed by the Limited Attentional Capacity Model, the ability of writing to approximate speech without some of the pressures of online planning, and provision of the opportunity to use precise language and syntax.

Why writing might better facilitate oral production

Writing has been touted as a unique way to enhance learning since the 1970s (Emig, 1977). Cumming (1990) posited that writing elicits attention to form and encourages learners to refine their expression to increase accuracy. Other reasons have been suggested to explain the way in which writing enhances learning, including that writing is a form of learning, approximates human speech, and that it supports learning strategies (Bangert-Drowns et al., 2004). The writing-to-learn perspective itself also sees writing as a valuable vehicle for learning (Harklau, 2002; Manchón, 2009; Manchón, 2011a).

It was hypothesized that participants in the written rehearsal (WR) group would focus on form (Williams, 2012) and have higher average measures of syntactic complexity while the oral rehearsal group (OR) participants would target, and have higher averages in, the realm of fluency.

Despite the paucity of prior studies concerning the role of writing in second language learning Williams (2012) speculated that writing could facilitate language development by providing the necessity and opportunity for students to use greater precision in language. This need for greater precision may encourage learners to make use of their explicit linguistic knowledge in planning and reviewing their production (Williams, 2012).

Blake (2009) examined face-to-face and text-based internet chats class formats and discovered that the written text-based chat group improved their oral fluency even more than the face-to-face discussion group. Although there are other reasons this format increased fluency gains (more frequent turn-taking, written corrective feedback, etc), Blake stated that the written text-based chat could help build oral fluency by facilitating automatization of lexical and grammatical knowledge. Indeed, greater automatization and accuracy in writing could lead to more accurate production, and therefore, when coupled with oral production, more fluent speech.

Harklau (2002) advocated for giving writing a more prominent role in second language acquisition as speaking is not the only communicative modality, and writing has been neglected over the years. Harklau noted that we should not neglect the examination of how students learn a language through writing.

Weissberg's study (Weissberg, 2000) has a unique connection to the current study as his participants were also native speakers of Spanish. He studied morpho-syntactic elements elicited with oral and written tasks and found that writing was generally the preferred medium for the emergence of new forms and the development of grammatical accuracy. He also stated that the relationship of L2 writing development with the simultaneous acquisition of L2 oral skills has received relatively little attention. In fact, some researchers have argued that oral communication is the basis of writing, represented by the "conversation to composition" idea (Berreiter & Scardamalia, 1982). However, Williams (2012) addressed this idea by saying that writing has often been seen as the result of acquisition, not as a facilitating factor of second language development. The current study examines the potential role of writing in facilitating language production, in accordance with the beliefs of Williams, Polio, and other writing-to-learn researchers. Manchón (2011a) in particular, discusses writing to learn language. This writing-to-

learn perspective has stimulated researchers to ask what specifically about the output created in the written modality can facilitate L2 development.

One of the only studies to examine if writing practice might facilitate oral production is Chau (2014). In his study, Chau (2014) examined whether planning with writing would enhance the fluency, complexity, and accuracy of L2 oral narratives. He used a picture-based narrative with three groups (no planning, planning without writing, and planning with writing) and measured the complexity, accuracy, and fluency of their subsequent oral narratives. He found that both planning groups performed better than the no-planning group. The planning with writing group was the most fluent of the three (as measured by syllables per minute), paused less in the middle of clauses, had significantly fewer lexical errors, demonstrated increased lexical variety (type-token ratio) while the measure of complexity affected in the planning without writing groups was the number of words per clause. In general, he found that planning, with and without writing, positively impacted L2 oral fluency though planning with writing had the biggest impact on speaking speed and lack of mid-clause pauses. This was a key finding as native speakers tend to also pause at the end of clauses rather than at mid-clause (Kahng, 2018; Skehan, 2009; Tavakoli, 2011; Skehan & Foster, 2005). Chau's study contributes to previous research on the benefit of planning on narrators' oral fluency (Mehnert, 1998; Crookes, 1989; Foster & Skehan, 1996; Gilabert, 2007; Tavokoli & Skehan, 2005; Yuan & Ellis, 2003; Wigglesworth, 1997).

The Story Continuation Task (SCT)

The technique chosen in the present study to elicit oral narratives was the story continuation task. The story continuation task (SCT) provides learners with an incomplete story text, which they are required to continue and complete in a coherent, logical story ending (Jiang,

2015; Peng, Wang, & Lu, 2018). The story used in the present study (from Wang & Qi, 2013) has an abrupt ending selected to intrigue and motivate participants. The SCT offers the benefit of providing grammatical structures and lexical items that could aid in student story continuation (Wang & Wang, 2015; Ye & Ren, 2019). Additionally, the SCT encourages alignment with the language and content of the story text so as to make learner narratives more coherent (Wang & Wang, 2015; Jiang, 2015). Thus far, the SCT has mostly been used to examine the effects of alignment in terms of lexical items used, errors committed, and the production of more complex writing in both assessment situations and task-based environments (Wang, 2012, 2015; Peng, Wang, Lu, 2018; Wang & Wang, 2014; Ye & Ren, 2019; Zhicheng & Lin, 2017).

It was anticipated that the written practice narratives would allow for the transfer of skills from writing to speaking, and based on Wang and Wang (2015), I assumed students would use some of the language from the story in their narratives. Thus, this might make the SCT an effective technique for providing students with language and ascertaining which modality will help them utilize this input more effectively.

In addition, the monologic oral narrative was chosen as a justifiable bridge between writing and speaking and has been used in multiple studies (Yuan & Ellis, 2003; Chau, 2014).

Focus on Complexity, Accuracy, and Fluency

Assessing the oral narratives after oral or written rehearsal is not straightforward, but I have chosen to focus on syntactic and lexical complexity, accuracy, and fluency as is customary in studies of manipulation of task features and conditions. As discussed earlier, in planning studies, the effects of planning have been assessed by focusing on these constructs. In addition,

planning may help learners focus on form with regard to grammar or the lexicon, or it may result in more fluent language because of easier access to grammar or vocabulary.

Complexity refers to the range of forms which appear and their degree of sophistication (Ortega, 2003). This study included measures that focuses on syntactic complexity; namely the mean T-unit length, variety of verb forms, and measures of lexical complexity such as lexical diversity and density. Multiple measures were selected in each category as per recommendation of Johnson's 2017 metanalysis, which found that many studies rely on only a few measures of complexity, accuracy, or fluency (Johnson, 2017). Although Johnson's metanalysis focused on CAF in L2 writing, many researchers and practitioners believe language proficiency to be multifaceted yet captured well with the dimensions of complexity, accuracy, and fluency (Ellis, 2003, 2008; Ellis & Barkhuizen, 2005; Skehan, 1998). As Housen and Kuiken (2009) state, CAF measures have been used as performance descriptors for oral and written assessment, for measuring progress, as well as for serving as indicators of learner proficiency (Housen & Kuiken, 2009).

T-Units as a Measure of Analysis

The unit of analysis chosen was the T-unit. Although the AS-unit is often used for speech, the T-unit is still a popular unit of analysis for written and spoken data (Foster, Tonkyn, Wigglesworth, 2000). As the speech produced in the current task had many similarities to writing (i.e., a narrative) the T-unit was chosen as the unit of analysis.

The T-unit was originally developed by Hunt (1965) as a more accurate way to measure syntactic development in children, by counting words per T-unit rather than words per sentence.

T-unit stands for minimally terminable unit and consists of a main clause and the dependent clauses that go with it (Bardovi-Harlig, 1992). T-units have been used in analyzing both written and oral production (Beebe, 1983; Larsen-Freeman, 1983) which enabled their use as a bridge between speaking and writing analysis. The present study analyzed the mean length of T-unit, achieved by dividing the number of words by the number of T-units in each narrative (Kawauchi, 2005; Mochizuki & Ortega, 2008; Chau, 2014). Mean length of T-unit has been shown to increase slightly in specific populations over time periods of at least 3 months in an ESL setting (Ortega, 2003). Research has shown a positive relationship between proficiency in an L2 and syntactic complexity as demonstrated by longer production units, such as the T-unit (Cumming et al., 2005; Lu, 2010; Ortega, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998).

Another way complexity was operationalized in the present study was by examining structural variety in the diversity of verb forms in terms of tense, aspect, modality, and voice (Ellis & Yuan, 2005; Foster & Skehan, 1996, Chau, 2014). Tensed verb forms and their level of variety is a valid measure for the present study as it seeks to examine the range of verbs used.

The calculation of lexical density was an examination of the ratio of content, or lexical, words per total words (Laufer, 1991; Mehnert, 1998), which was automatically calculated using the online lexical profiling software https://www.lextutor.ca/vp/eng/ (Laufer & Nation, 1995; Meara & Fitzpatrick, 2000).

The measure of lexical diversity (MTLD) was chosen as the measure for lexical diversity as it has been shown to not vary according to text length (Mccarthy & Jarvis, 2010; McCarthy, 2005). This was an important choice as some texts were as short as 62 words and others as long as 470 words and the comparison needed to take varying lengths into account.

Accuracy

The written performance measures for accuracy listed by Wolfe-Quintero et al (1998) are often used for assessing oral performance as well (Levkina & Gilabert 2012; Michel, 2011). Polio and Shea's (2014) study brought light to the fact that no measure of written accuracy for writing has, as of now, jumped out as best, especially considering none of the error measures in their study changed over time. Though researchers would like to develop a universal index of accuracy, it is not clear if this is possible (Polio & Shea, 2014). Some researchers believe that general error density, including measures such as error-free units are sensitive to predicting differences in experimental conditions (Foster & Skehan 1999; Iwashita et al., 2008). Researchers, including Tonkyn (2012), have used accuracy measures such as error-free units, syntactic errors, or lexical errors for oral language. Although such measures may accurately predict differences, at the same time it is important to keep in mind that errors may increase (decreasing accuracy) as complexity increases (Skehan & Foster, 1997). Rather than counting each individual error type per unit, or coding everything to be analyzed according to target-like use (Ellis & Barkhuizen, 2005), which can be challenging to achieve inter-rater reliability on, accuracy was examined in light of total errors and subdivisions of three error types. Errors in morphology, syntax, and prepositions were examined and added together and transformed to yield errors per 100 words in order to make direct comparisons. This allowed insight into the specific error sub-categories while also being easier to categorize and achieve inter-rater reliability on.

Fluency

Fluency was one of the three overarching measures chosen which draws largely on the oral fluency and pausology research versus the predominantly writing-centered research measures

offered for complexity and accuracy. The story continuation task (SCT) bridges speaking and writing as it is an integrated task (Ye & Ren, 2019; Zeng, Mao, & Jiang, 2017) composing the reading and comprehension of a text with the writing and final oral production.

Although fluency is difficult to define (Kormos & Denes, 2014) certain aspects of speech have been shown to have high correlations with the perception of fluency. Silent pause rate *within* a clause has been proven to have the strongest correlation with L2 fluency ratings and perceived fluency is greatly influenced by pause location: within-clause pauses lower fluency ratings more than between-clause pauses (Kahng, 2018). It is believed that within-pause clauses reflect L2 speakers reduced cognitive fluency outwardly to those listening, yet personality traits and speaking style have been found to influence average pausing duration (De Jong et al., 2013). Fluency is of paramount importance because everyone from teachers, assessors, listeners, to the students themselves considers it to be important (Schmidt, 2000). Speakers are already less fluent in their L2 (Segalowitz, 2010) and need to find ways to reduce this lack of fluency in order to improve overall speaking skills and performance on proficiency tests and assessments (Cucchiarini, Strik, & Boyes, 2002; Housen, Kuiken &Vedder, 2012; Iwashita, Brown, McNamara, & O'Hagan, 2008).

Fluency was operationalized in three ways; speed fluency or speech rate, repair fluency, and pause length/location (Housen & Kuiken, 2009; Skehan, 2009; Chau, 2014). Speed fluency was measured with syllables per minute (Kormos, 2014; Yuan & Ellis, 2003; O'Brien, Segalowitz, Freed, & Collentine, 2007) which was measured online using the Poetry Soup Syllable Counter https://www.poetrysoup.com/syllables/syllable_counter.aspx. Repair fluency was measured as the number of reformulations, self-corrections, repetitions, replacements, and false-starts per minute (Foster & Skehan,1996; Kormos, 2014; Skehan & Foster, 1999; Tavakoli & Skehan,

2005). To calculate this numerically the number of dysfluency phenomena was divided by total speaking time measured in seconds as in Elder and Iwashita's (2005) study and then multiplied by 60 to obtain the number per minute.

Pause lengths were measured at 1+ second for clause junctures and .39 miliseconds or more in mid-clause locations as this has been shown to be a more critical marker of fluency than pauses appearing in positions at clause junctures (Kahng, 2014, 2017). Silent and filled pauses were counted separately and only silent pauses had their location noted, as appearing in mid-clause or mid-reformulation/filler position.

Complexity and accuracy, coupled with the last measure, fluency, are viewed as basic dimensions of L2 performance and development (Housen & Kuiken, 2009, Larsen-Freeman, 2006; Skehan, 2003; Wolfe-Quintero et al., 1998). For this reason, all three measures have been considered in an attempt to visualize the potential differential impacts of writing versus speaking practice in correlation to a final monologic oral task.

The present study intends to combine what is known about using complexity, accuracy, and fluency measures to obtain information about linguistic growth and differences in the analysis of a final oral story continuation task performance in an EFL setting and to analyze differences in speaking and writing used as rehearsal/practice before the task itself. Based on what is known about pre-task and online planning, two groups are implemented: the writing practice group, which is highly pre-task planning oriented, and the speaking practice group, which contains elements of pre-task and online planning as well as a certain degree of task repetition. The research question is as follows:

Research Question

1. What is the impact of writing rehearsal, or practice, in comparison to speaking rehearsal when preceding an L2 oral narrative task, with regard to complexity, accuracy, and fluency measures?

It is hypothesized that the participants in the writing rehearsal/practice group (WR) will have fewer form-based errors and a more advanced level of vocabulary while the speaking rehearsal/ practice group (OR) will have increased student fluency and speaking speed but grammatical structures and vocabulary which are less complex.

METHODS/RESEARCH DESIGN

This study is a between subject design with two conditions: rehearsal (planning-practice) by writing (WR) or rehearsal by individual speaking (OR). Groups of participants taken from

novice and advanced classrooms were randomly divided into two treatment groups and asked to continue and finish a story with the ending removed. There was an equal mix of advanced and beginner level L2 speakers in each treatment group. The specific directions given by the researcher to students in the WR group were provided orally and written on their handout. The directions congratulated the students on reading the story and told them their job was to finish it. They were told they would have approximately 20 minutes to write out how they wanted the story to end and that they could make an outline or brainstorm before writing if they liked. Many students began writing immediately without engaging in any outlining. They were also informed that, "the goal of this writing practice is to help you prepare for telling your final version out loud". The directions for the OR group differed only in that they were told they would have approximately 20 minutes to practice telling their story ending out loud to themselves. They were informed they could record with their phone or Whatsapp but "the goal is to just practice for your final version." See Appendix 1: *Story and Directions for Groups* for a copy of the handout the participants were given.

Participants

At the time of data collection, participants were EFL teachers with a common L1 of Spanish participating in the English language teaching (ELT) professional development program where the researcher and a team of other ELT professionals were giving a professional development (PD) course. The participants in the study included 34 Mexican teachers of English as a foreign language in the state of Tabasco, Mexico. The teachers varied in terms of teaching setting (i.e., university, private business, or vocational college), years of experience teaching English, exposure to the language, and proficiency level. There was an equal mix of L2 English

speakers included in each group with 9 advanced and 8 beginners in each group. All participants partook in an intensive summer ELT course hosted at a university in southern Mexico.

Prior to enrollment in the course, teachers were given an informal oral interview test of English language ability, similar to the IELTS speaking test, in order to arrange classes according to participant level. The oral-interview styled tests consisted of personal questions eliciting different tense use and were conducted by one to two native-speaking professional teachers of English with experience in ESL and EFL instructional settings, using a rubric to guide judgment of oral proficiency level. Some of the participants had never conversed with a native English-speaker before but had studied teaching, linguistics, or English in a post-secondary setting. Gender distribution was nearly even across both groups with 11 females and 6 males in the WR group and 12 females and 5 males in the OR group. The experiment was conducted after 1.5 weeks of intensive PD sessions given exclusively in English had already transpired.

Materials

Each student received a print-out of the story *Park Avenue Surprises* (Wang & Qi, 2013). The story was 329 words and the students were given approximately 5 minutes to read it. See Appendix I for the story and the directions each group received. This story continuation task was chosen for multiple reasons. Firstly, a story-continuation writing task requires students to comprehend a reading passage and cohesively and creatively extend that story to complete it in a sensible way (Ye & Ren, 2019). Secondly, the story had been successfully used in multiple studies (Wang & Qi, 2013; Ye & Ren, 2019). Lastly, the story was cut off abruptly so as to be intriguing to the participants and increase their engagement with the task. The content of the

16

story was determined to be linguistically and culturally accessible from pre-experiment discussions with locals.

The reading level, and resulting cognitive demand presented by the story, was analyzed using the Flesch-Kincaid scale on readability and determined to be 2.9, meaning that it was easy to read (Ye & Ren, 2019). Due to the low lexical level of the story it is unlikely that even the lower-proficiency participants encountered difficulties in comprehension. However, when it came to producing the language connected to a bank robbery and false accusations some students struggled to find the correct vocabulary and a number of false cognates, or borrowed words from Spanish, were seen.

After the experiment audio files were uploaded, transcribed, and converted to WAV files, in order to be analyzed with PRAAT.

Procedure

The experiment was conducted during class time in the classrooms, starting with the least-proficient level and ending with the most advanced classroom. During the experiment, the researcher went to the participants' classroom, explained the tasks, that participation was entirely voluntary, obtained signed consent forms, and provided contact information for herself via telephone and WhatsApp. Within their classroom participants were randomly assigned to one of the two experimental groups with a total of 17 in the WR group and 17 in the OR group. Then a chart was drawn on the board, including basic instructions (pictured below in Figure 1.), and the researcher explained, in basic English, the steps each student would go through.

Step 1: Read the story fragment (5 minutes)

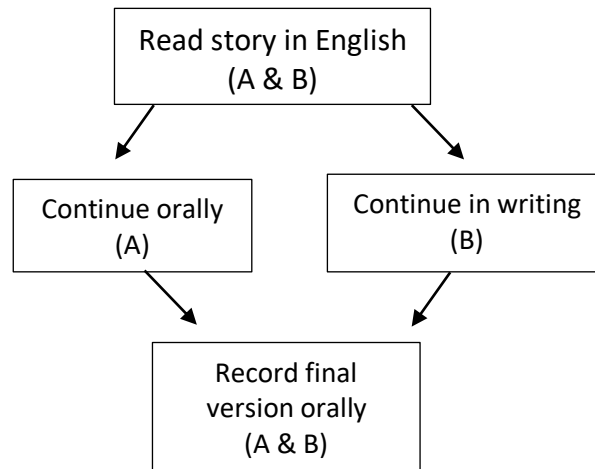*Both groups read short story with ending removed in English (approximately 329 words)*

Step 2: **Create an ending for the story** *(20 minutes)*

Group A: orally (can re-do and practice as much as desired)

Group B: in a written format (rewrite or reformat as much as desired)

Step 3: **Orally record their final version of the story (5 minutes)**

Figure 1.
Experiment Instructions

As the story was just over 300 words, the students were given approximately 5 minutes to read it. After 5 minutes the researcher had the WR group remain in the classroom for their 20 minutes of writing practice while the OR group left the classroom and dispersed themselves to orally rehearse how they planned to finish the story.

Some students in the OR group (monitored by the researcher) chose to orally brainstorm, practice the ending and critique themselves, record their ideas into WhatsApp, or speak out multiple versions of an ending and choose the one they most enjoyed. Observations from inside the classroom with the WR group were that some students immediately started writing, brainstormed ideas, or re-read the story, while others took a moment to think before writing. No participants were allowed access to outside sources. The WR practice was collected at the end of

the session while only a small percentage of OR elected to record and send their spoken practice via WhatsApp.

After the allotted planning time was up all students audio-recorded their story ending and sent it to the researcher via WhatsApp.

Measures

Measures of complexity, accuracy, and fluency were used to analyze the final, oral story recordings. Table 1 includes a list of measures; note that reliability has not been measured at this time. Regarding complexity, the oral transcriptions were divided into T-units (Hunt, 1968) to allow further analysis.

Table 1.
CAF Measures Used

Measure	Description
T- unit length	Average number of words per T-unit
Average word length	Average number of letters per word.
Verb form variety	Different verb tenses/aspects used
Lexical diversity	Calculated using MTLD
Lexical density	Content words out of total words
Morphologic errors/ 100 words	Incorrect word forms, articles, subject-verb agreement
Preposition errors/100 words	Any extra, missing, incorrect prepositions
Syntactic errors/100 words	Incorrect word order, extra/missing verbs or subjects
Total errors/ 100 words	Morphology/preposition/syntax errors combined
Syllables/ minute	Total syllables divided by length, multiplied by 60
Reformulations/ minute	Repetitions, self-corrections, false-starts, replacements
Filled pauses/ minutes	Pauses with fillers, such as *uhm, ahh, uhh*
Silent pauses/ minute	Pauses without fillers, pure silence
Mid-reformulation pauses/ minute	Silent pauses occurring in the middle of a reformulation
Mid-clause pauses/minute	Silent pauses occurring in the middle of a clause

Complexity

Factors considered for complexity were verb form variety, (i.e. how many different verb forms are used, as shown by tense, aspect, voice, modality; see Chau, 2014), lexical diversity, as

19

measured by MTLD (McCarthy & Jarvis, 2010), t-unit length, average word length, and lexical density (content words/total words as calculated Lextutor.ca Vocab Profile.

Accuracy

Accuracy was examined considering three error types, following Polio and Shea (2014). Errors in morphology, syntax, and prepositions were examined and added together and transformed to yield errors per 100 words in order to make direct comparisons between participants and groups. Morphological errors included incorrect word forms, lack of subject-verb agreement, non- target-like use of articles, wrong pronouns, verb form problems such as incorrect tense, aspect, voice, or missing infinitives or modals. Syntactic errors included incorrect word order, missing constituents, extra verbs or subjects, improper use of relative clause pronouns (who/that/ which/etc). Prepositional errors were defined as any extra, incorrect, or missing prepositions.

Fluency

Fluency was operationalized in three ways; speed fluency, repair fluency, and pause length/location. Speed fluency was measured with syllables per minute, while repair fluency was the number of reformulations, self-corrections, repetitions, replacements, and false-starts per minute. Pause lengths were measured at 1+ second for clause junctures (Kahng, 2014) and .39 milliseconds or more in mid-clause locations (Kahng, 2014). Silent and filled pauses were counted separately and only silent pauses – not occurring at clause junctures were considered and calculated as appearing mid-clause or mid-reformulation.

Analysis

Analysis was conducted using SPSS Mann-Whitney, 2-independent variable non-parametric tests as data did not follow normal distributional patterns. Effect size was calculated

using coefficient, *r* as both groups had different standard deviation. The formula used in calculations divided Z by the square root of N ($r = Z / \sqrt{N}$). SPSS and Excel were utilized to produce data visualizations and run various analyses. The reliability of the complexity and accuracy measures will be calculated using a second rater to ascertain inter-rater reliability.

RESULTS

In order to answer the research question, I first ran and reviewed descriptive statistics followed by inferential nonparametric tests. It was hypothesized that the participants in the WR group would have fewer form-based errors (accuracy) and a more advanced level of vocabulary (higher average complexity) and possibly lower scores in fluency measures than the OR group, while the OR participants would have increased fluency and speaking speed but less accurate grammar and complex vocabulary. The results revealed a more complex picture than was anticipated.

Descriptive Statistics

The first three columns in Table 2 show three basic metrics for the writing and speaking practice groups: mean length, in seconds; pruned narrative word count; total T-units. These were recorded not as CAF measures, but as components needed to calculate other metrics. However, the speaking group did have narratives that were almost 50 seconds longer and over 30 words longer. Time on task came into play here as both groups had 20 minutes to rehearse but writing takes longer, and so participants may not have been able to entirely complete their written narrative before the final task performance. The OR group may also have benefitted from more direct task-repetition as the rehearsal was in the same modality as the final task performance.

The next five columns in Table 2 show the five measures chosen for complexity. For measures of complexity the lexical diversity, as measured by MTLD (WR: 54 and OR: 43), shows a noticeable difference between groups.

Table 2.
Complexity Measures

		Length (secs)	Length (word count)	Length (T-units)	C:T-unit length	C:verb variety (tense)	C:MTLD lexical diversity	C:avg word length	C:lexical density
WR	*M*	67.1	157.5	14.3	11.4	3.2	54.2	4.0	0.4
	N	17.0	17.0	17.0	17.0	17.0	17.0	17.0	17.0
	SD	26.9	64.3	5.7	2.7	1.1	14.9	0.3	0.0
OR	*M*	116.8	189.5	18.7	10.0	3.4	43.1	4.0	0.4
	N	17.0	17.0	17.0	17.0	17.0	17.0	17.0	17.0
	SD	74.4	124.2	10.7	1.8	1.4	7.3	0.2	0.0
Tot	*M*	91.9	173.5	16.5	10.7	3.3	48.7	4.0	0.4
	N	34.0	34.0	34.0	34.0	34.0	34.0	34.0	34.0
	SD	60.6	98.7	8.7	2.3	1.2	12.9	0.2	0.0

Regarding accuracy, the WR group did have fewer errors per words in all categories (though the descriptive statistics were not tested for effect size). The four accuracy measures are displayed in Table 3 and are not statistically significant.

Table 3.
Accuracy Measures

		A:errors/ 100	A:morph errors/100	A:prepo errors/100	A:syntax errors/100
Writing rehearsal	*M*	**4.7**	**3.3**	**0.9**	**0.5**
	N	17.0	17.0	17.0	17.0
	SD	2.3	2.4	0.8	0.5
Oral rehearsal	*M*	**6.7**	**4.7**	**1.3**	**0.7**
	N	17.0	17.0	17.0	17.0
	SD	4.1	2.8	0.7	1.1
Total	*M*	**5.7**	**4.0**	**1.1**	**0.6**
	N	34.0	34.0	34.0	34.0
	SD	3.4	2.6	0.8	0.8

Displayed in Table 4 you see the six fluency measures. For fluency, the WR group has noticeably better measures in syllables per minute (174 vs 149), half as many filled pauses per minute, 25% fewer reformulations per minute, fewer silent pauses in the middle of clauses (4.5 vs 5.6), and also fewer silent pauses occurring in the middle of a reformulation (1.8 vs 2.9). The descriptive statistics appear first and are followed by the inferential tests.

Table 4.
Fluency Measures

		F:syllables (/min)	F:filled pause (/min)	F:total silent pauses (/min)	F:reforms (/min)	F:mid-clause pause (/min)	F:midreform pauses (/min)
Writing	M	173.8	0.5	10.6	2.0	4.5	1.8
	N	17.0	17.0	17.0	17.0	17.0	17.0
	SD	27.5	1.0	6.5	2.1	3.2	1.8
Oral	M	125.0	1.4	13.8	3.2	6.8	4.1
	N	17.0	17.0	17.0	17.0	17.0	17.0
	SD	34.1	1.9	6.4	3.1	4.9	3.3
Total	M	149.4	1.0	12.2	2.6	5.6	2.9
	N	34.0	34.0	34.0	34.0	34.0	34.0
	SD	39.3	1.6	6.5	2.7	4.2	2.9

Inferential Statistics and Tests of Significance

For inferential statistics Mann-Whitney non-parametric, 2-variable tests were chosen, along with coefficient *r* for effect size. These tests were chosen due to the abnormal data distribution, potentially due to the inclusion of such a wide variety of levels. Two overall fluency measures and one overall complexity measure showed statistical significance, while three other fluency measures approached statistical significance.

Mann-Whitney nonparametric test data is presented in Table 4. The results indicate that the WR outperformed the OR group in the 2 measures of fluency and one of complexity (lexical diversity). Two effect-sizes were classified as medium effect (lexical diversity, silent pauses in mid-reformulation) and one with a large effect size (fluency measured in syllables per minute).

For measures of accuracy no statistically significant difference was shown, and only lexical diversity from complexity measures showed a significant difference. The WR group outperformed the OR group in several measures of oral fluency.

Table 5 shows the results of the Mann-Whitney test, with the three statistically significant results bolded and effect sizes calculated for the three significant findings, which are lexical

diversity (MTLD), syllables per minute, and silent pauses occurring in the middle of

reformulations.

Table 5.
Ranks

	Group	Mean Rank	Sum of Ranks	Z	Sig. (2-tailed)	Effect size
Word count	WR	17.12	291.00	-0.224	0.823	
	OR	17.88	304.00			
Length (sec)	WR	14.18	241.00	-1.949	0.051	
	OR	20.82	354.00			
Total T-units	WR	15.74	267.50	-1.035	0.301	
	OR	19.26	327.50			
T-unit length	WR	20.00	340.00	-1.464	0.143	
	OR	15.00	255.00			
Avg. word length	WR	18.26	310.50	-0.450	0.653	
	OR	16.74	284.50			
Verb variety	WR	17.15	291.50	-0.215	0.830	
	OR	17.85	303.50			
MTLD	WR	21.41	364.00	-2.290	**0.022**	**-0.393**
	OR	13.59	231.00			
Lexical Density	WR	19.29	328.00	-1.057	0.291	
	OR	15.71	267.00			
Total Errors /100	WR	15.32	260.50	-1.275	0.202	
	OR	19.68	334.50			
Morph. Errors /100	WR	15.12	257.00	-1.396	0.163	
	OR	19.88	338.00			
Syntax errors /100	WR	17.47	297.00	-0.018	0.986	
	OR	17.53	298.00			
Prep errors /100	WR	15.26	259.50	-1.315	0.189	
	OR	19.74	335.50			
Syllables /min	WR	23.53	400.00	-3.532	**0.000**	**-0.606**
	OR	11.47	195.00			
Reformulations /min	WR	15.15	257.50	-1.391	0.164	
	OR	19.85	337.50			
Filled pauses /min	WR	15.18	258.00	-1.620	0.105	
	OR	19.82	337.00			
Silent pauses /min	WR	14.74	250.50	-1.524	0.127	
	OR	20.26	344.50			
Mid-reform. pauses /min	WR	13.65	232.00	-2.271	**0.023**	**-0.39**
	OR	21.35	363.00			
Mid-clause pauses /min	WR	14.71	250.00	-1.637	0.102	
	OR	20.29	345.00			

To see if participants in the WR group had fewer grammatical errors we must examine the

results for accuracy. Overall accuracy measures were not significant, with morphological errors

per 100 words coming in at .163, preposition errors per 100 words at .189, syntactic errors per

100 words at .986, and total error count per 100 words at .202, using Mann-Whitney non-parametric, 2 independent variables, 2-tailed significance. Despite lack of statistical significance there was a definite variance in mean rank. The WR group had a mean rank of 15.32 for errors per 100 words compared to 19.68 in the OR group; 15.12 morphological errors per 100 words in the WR group versus 19.88 in the OR group; and the WR group had a mean rank of 15.26 preposition errors per 100 words compared to 19.74 for the OR group. Out of complexity, accuracy, and fluency it was accuracy that was found to be the least statistically significant and least influenced by which group the participants were in.

For complexity, the measure found to be statistically significant was lexical diversity, as measured by MTLD with .022 significance and a Z score of -2.290 on the Mann-Whitney test. There was a medium effect size as calculated with the coefficient *r*, -0.393. It was the WR group which had a higher lexical diversity score. In fact, the WR group had a mean rank of 21.41 compared to only 13.59 for the OR group. Other measures investigated, such as verb form variety, average word length, lexical density (as measured by content words/100 words), and T-unit length were nearly the same in both groups and appeared to not distinguish complexity between WR and OR in this study. WR participants used a greater range of words and thus garnered the higher lexical diversity score, distinguishing their final task in one measure of complexity.

To answer whether the OR group had higher fluency scores than the WR group, we must look at the two fluency indicators shown to be statistically significant in this study. Fluency speed, or speech rate, as measured by syllables per minute was statistically significant at .000 with a large effect size of -0.606. The group to have a mean rank double to the other was not, as anticipated, the OR group. Rather, the WR group averaged 23 syllables per minute compared to

the OR group's 11. Additionally, the silent pauses occurring in the middle of a reformulation, repetition, or verbal filler (***Uhm** After that **eh** John feel petrified [**PAUSE**] because [**PAUSE**] **uhm** and he didn't know what to do*) were statistically significant at .023 with a medium effect size of -0.390. The mid-reformulation pauses per minute in the WR group had a mean rank of 13.65 compared to 21.35 in the OR group.

Lastly, I would like to answer, according to this study, what the impact of writing rehearsal is, when used as a type of practice preceding an oral production task for EFL learners. The WR group performed better than the OR group on all measures that were, or approached, statistical significance, including and in particular, fluency. In this study writing was shown to double the speech rate (evidenced in syllables per minute), highly increase the lexical diversity, and also significantly decrease the silent pauses occurring in the middle of reformulations and filled pauses. Total silent pauses occurring in the middle of the clause had a significance of .003, though that was lessened to .102 when standardized to silent mid-clause pauses per minute. Several other measures approached some significance and there was noticeable difference in many mean ranks, even if they are not seen as statistically significant.

Considering the effect sizes and significance of multiple measures it can be said that, for this study and the demographic of Mexican EFL teachers with low and advanced English levels, writing has a positive impact on speaking tasks with no noticeable negative effects. A combination of pre-task writing, and online speaking practice before the final oral task could yield even better results for students.

DISCUSSION

The present study adds to the yet under-investigated area of how the modality of writing contributes to oral performance by showing that allowing L2 learners to practice and plan via writing can be effective for improving oral performance. The results show that WR outperformed the OR group on 3 measures: lexical diversity, syllables per minute, and silent pauses occurring in the middle of a reformulation. The WR group had noticeably higher scores than the OR group in terms of fluency, but complexity scores showed a smaller variation between the groups while accuracy scores displayed little difference and no significant measures. The results support previous research, such as Chau (2014) which indicated that groups planning with writing are more fluent, pause less in unnatural locations. Linguistic improvement does occur when employing the written modality, specifically, creating gains in the area of fluency.

The finding that no accuracy measures were statistically significant and lexical diversity was the sole significant measure for complexity may indicate that though the WR group may not have been able to recall the accurate verb form or long clause they wrote but they were able to speak more fluently. Though the OR group did have longer narratives concerning total seconds and word count, the WR group narratives had more syllables per minute, showing they were speaking faster.

Although the findings in the area of fluency were not completely expected, the Yuan and Ellis (2003) study on the effects of online and pre-task planning go a long way towards explaining this result. To reiterate a key finding from their study; online (in-the moment) planning enables participants to give more attention to grammatical accuracy but results in reduced fluency and induces reliance on more basic vocabulary. Pre-task planning encourages attention to conveying the message with greater fluency and lexical variety.

28

In the present study both WR and OR group participants were given 20-minutes of rehearsal time before the final task. WR participants used their time to write and plan out their story (mainly pre-task planning). However, they were not orally producing it or monitoring their speech, whereas OR students had access to limited online planning before the final performance as they were allowed to practice their story ending orally as many times as they liked during their time.

When the time was up, the WR group had overall well-developed plot lines and had chosen appropriate lexical items to convey their meaning while the OR group had not. Thus, it was the WR group who spoke with greater fluency and lexical variety, just as Yuan and Ellis' (2003) study found. A larger sample size using participants all from the same level could magnify and clarify findings. The differences could also be attributed to the mix of advanced and novice-low English learners, dispositions, or comfort with speaking tasks under a minor time pressure.

The results show the WR group increased in speech rate, decreased in mid-reformulation pausing, and increased in lexical diversity, so we can conclude that given planning learners engage in cognitive activities which lead to selecting appropriate vocabulary and other means to fluently convey their message.

There are some limitations of the present study that should be acknowledged. Firstly, that the oral practice group data collected was an extremely small percentage of the total participants (three out of seventeen in the OR group recorded and sent their practice). Although the researcher circulated the area WR participants were practicing in and generally noted what different participants were doing no written notes were taken so as to avoid influencing natural task performance. However, the written practice sheets were collected from the WR group

participants. Since only a small number of participants from the OR group recorded and sent their data the rehearsal data is not getting analyzed.

A study employing multiple story-continuation tasks given several days apart would expand our knowledge of how learners use writing to improve their speaking. Think-alouds, or similar methods to obtain participant thoughts, were not conducted but could give insight into the mental processes and choices of participants in the writing and speaking practice groups.

CONCLUSION

This study investigating L2 teachers of English in an EFL context and the impact of pre-task writing rehearsal opportunities versus pre-task speaking rehearsal has demonstrated that writing before an oral performance task is a highly desirable practice modality. As there are not many studies examining the impact of writing on speaking it is difficult to discuss the extent to which the results obtained can be generalized to other learners in different contexts. Generalizability remains to be researched and studies should be conducted with larger groups of participants ($n=34$) of a more homogenous English level (16 advanced, 18 low-mid). One should use caution when seeking to generalize the results to other contexts as each location and group of learners brings unique variance. Though the results obtained from the WR group are significant and have medium to high effect sizes in three subcategories (stemming from fluency and complexity), take care to not over-interpret the study.

Nevertheless, results found in this study shed light on the versatility of writing in developing both complexity and fluency related to the use of writing and speaking practice before oral task performance. Previous studies have brought the impact of pre-task and online planning to our attention, focused on writing-only or speaking-only studies, yet few have combined the two. This study suggests that a key factor in increasing fluency may be whether the learners have the option to plan via the written modality. In future studies of planning it will be necessary to carefully consider and regulate conditions under which tasks are performed to control on-line planning. There is general agreement that writing is a helpful skill to develop but there is a lack of knowledge about the impact writing can have on speaking.

Finally, I will consider implications of this study for language pedagogy. The use of SCT in the classroom can promote student motivation, fluency, complexity, and possibly encourage

31

higher accuracy. Teachers can facilitate and incorporate a SCT by giving students an interesting text of 250-500 words (dependent on student level, absolute beginners would need shorter texts) to read. A lexical profiling should be completed first and any challenging vocabulary pre-taught to students so as not to hinder comprehension of the story. Teachers may engage in small-group discussion after the story reading if concerned their students may not have understood accurately. After the story reading give students 10-20 minutes to write how they would like the story to end. This gives them time to create the concepts they want to convey and search for the necessary lexical items. After the time is up take away the written story from each student. At this point you may have students record their ending orally (as in this study), tell their story to another student (additional practice and task repetition), or practice their story orally and reformulate as many times as they would like (online planning). After the task is completed the teacher may choose to evaluate oral performances using measures of complexity, accuracy, fluency, as was done in this study. Post-task discussion is encouraged to seek feedback from students in challenges faced and reflection on any skills they learned or improved during the task.

APPENDICES

Appendix A: Story Handout from Wang and Qi (2013) and Directions for Groups

Park Avenue Surprises

        An unusual thing happened to John when he was on the way to work one day. As he walked along Park Avenue near the First National Bank, he heard the sound of someone trying to start a car. He tried again and again but couldn't get the car moving. John turned and looked inside at the face of a young man who looked worried. John stopped and asked, "It looks like you've got a problem," John said.

"I'm afraid so. I'm in a big hurry and I can't start my car."

"Is there something I can do to help? John asked. The young man looked at the two suitcases in the back seat and then said, "Thanks. If you're sure it wouldn't be too much trouble, you could help me get these suitcases into that taxi over there."

"No trouble at all. I'd be glad to help."

        The young man got out and took one of the suitcases from the back seat. After placing it on the ground, he turned to get the other one. Just as John picked up the first suitcase and started walking, he heard the long loud *noise* of an alarm.

It was from the bank. There had been a *robbery*!

Park Avenue had been quiet a moment before. Now the air was filled with the sound of the alarm and the shouts of people running from all directions. Cars stopped and the passengers joined the crowd in front of the bank. People asked each other, "What happened?" But everyone had a different answer.

        John, still carrying the suitcase, turned to look at the bank and walked right into the young woman in front of him. She looked at the suitcase and then at him. John was surprised. "Why is she looking at me like that?" He thought. "The suitcase! She thinks I'm the bank robber!"

John looked around at the crowd of people. He became frightened, and without another thought, he started to run…

Writing Practice:
Congratulations on reading the start of this story – your job is to finish the story. How does it end? What happens to John and the true thief who stole from the bank? You will have approximately 20 minutes to write out how you want the story to end. You may make an outline or brainstorm before writing if you would like. The goal of this writing practice is to help you prepare for telling your final version out loud.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Speaking Practice:
Congratulations on reading the start of this story – your job is to finish the story. How does it end? What happens to John and the true thief who stole from the bank? You will have approximately 20 minutes to practice your story ending out loud. You may record with your phone or Whatsapp but the goal is to just practice for your final version.

Speaking Logistics:
Use WhatsApp (can record up to 15 minutes)
When you hit the record button slide finger up to "lock in" the recording, and just hit send when done.

Appendix B: Coding Guidelines

Coding Guidelines for Complexity
1. **T-unit length (mean length of T-unit)**
   a. Divide the number of words by the number of T-units in each pruned narrative.
   b. Pruned means that fillers and dysfluencies were excluded from the total word count. Such as *ummmm, ah, hmm, the the the car*, and so on.
   c. A T-Unit is one main clause and all the subordinate clauses attached to it. Each t-unit is the shortest sentence that can be created while not being a fragment. (Hunt, 1965)
   d. Two days later a man was found.
   e. Nobody knew his name since no personal information couldn't be found on him but a dollar *hanging out of a small hole in a big suitcase and a happy face.
   f. *Wouldn't break here because a fragment would be created/an adjective clause.
   a. If the narrative had 168 words and 11 T-units the mean T-unit length is (168/11): 15.27 versus another with 113 total words and 16 T-units, having a mean T-unit length of 7.1 words.
2. **Variety of verb forms**
   a. Count total number of different forms used per narrative.
   b. Consider tense (past, present, future) aspects (simple, progressive, perfective).
   c. While he **was running** he **could see** many people pointing at him.
   d. Two days later a man **was found**.
   e. Indeed, it was a terrible mistake **to have helped** a stranger.
   f.
3. **Lexical Density: MTLD**
   a. automatically calculated using CohMetrix 3.0

Coding guidelines for accuracy measures
Exclude all fillers/dysfluencies (reformulation/self-correct, substitutive repetitions, replacements, false starts) from total word count first) = pruned narrative
1. **Number of errors/100 words**
   a. Calculate: Divide total number of errors by number of words in each narrative and then multiply the result by 100%
   b. Count consistently repeated errors only once

2. **Syntax Errors**
   a. **incorrect word order**
      i. *I didn't know what should I do*
   b. **missing constituents**
      i. He **was** well-known in the neighborhood and[THEY] would easily locate him.
   c. **improper uses of relative clause pronouns**
      i. (who/that/which/whose/where/when)
      ii. They could not finish their plan because the car where [WHICH/THAT] they are going to go away [IN] was broke.
3. **Morphology Errors**

- **a. Word Forms**
    - i. noun instead of adjective, verb, adverb, etc
    - ii. *He was surprising because of the people were running everywhere.*
    - iii. The police officers were not very justice (just).
- **b. Subject-verb agreement**
    - i. But he don't know *what's the problem.*
    - ii. He offered to help a young man that has a big problem.
- **c. target-like use of articles**
    - i. lack of obligatory article, article where none is required
    - ii. He was caught and sent to the jail.
    - iii. When he was in front of the bank he hear a sound of [A]siren.
    - iv. He changed his mind because he wasn't [THE/A] robber.
- **d. Wrong pronouns**
    - i. *gender/case*
    - ii. Well, one surprise, the man who helped he was with the police.
    - iii. He saw a young man having problem with his car because he didn't start.
    - iv. He started to run but he couldn't stop and he changed [HIS] mind
- **e. Verb form problems**
    - i. *tense-aspect, passive voice, missing/extra to-infinitives, modals*
    - ii. He left the suitcase on the street and **run** away.
    - iii. And the police catch the stoler**.**
    - iv. I noticed that he had robbered a bank
    - i. He decided to cross and to thrown himself into the water

4. **Preposition Errors**
    - a. all missing/extra/incorrect prepositions
    - b. John **looked** around [AT] the crowd of people
    - c. Just in that moment people arrive to him.
    - d. After two days somebody **called** ~~to~~ the police and **said** where John was.
    - e. he **ran** as fast [AS] he could.

Coding guidelines for fluency measures
*Using original narratives
1. **Speed fluency/Speech rate**
    - a. **Number of syllables per minute:**
    - b. *Divide total number of syllables by total number of seconds (including pause time once speaking has begun), and then multiply the result by 60.*
    - c. 225 syllables/79 seconds= 2.8481*60= **170** syllables per minute
2. **Breakdown fluency/Pauses**
    - a. **Number of (silent) pauses per minute:**
    - b. *Divide the total number of silent pauses by the total number of seconds and multiply by 60.* Use PRAAT to measure pause length while listening to audio and looking at complete transcript.
    - c. Count only pauses of .39 seconds midclause or mid-reformulation and 1+ second at clause junctions.
    - d. For pauses of 1+ second write [PAUSE]

e. For pauses of <1 second write [pause]

f. Calculate filled (ah, uhm, you know, well…) pauses and unfilled (silent) pauses separately.

g. Mark position of silent pauses if mid-clause/mid-reformulation.

h. *He became frightened and without another thought he started to run away[PAUSE] He left the suitcase on the street and run away with not[PAUSE] with none direction in particular. When he stopped because he couldn't run aw[pause] run anymore[PAUSE]*

3. Repair Fluency

  a. Number of dysfluencies (reformulation/self-correction, repetition, replacements, false-starts : see below) per minute. Divide the total number of dysfluencies by total number of seconds and multiply by 60.

  b. Repetitions: syllables, words, phrases, clauses immediately repeated with no modification. (verbatim)

    i. Who, **whooo**, eh**,** asked like maybe one, **one**, uh**, one** million of dollar.

  c. Reformulations/self-corrections are words, phrases, clauses repeated with some modification (to morphology, syntax, pronunciation, word order) Includes replacements (lexical items immediately substituted for another).

    i. And run away with **not**, with **none** direction in particular. When he stopped because he couldn't **run aw… run anymore.** He felt so stupid and *walked* to the to.. to **her**..**hus**… **to his** home.

  d. False starts are utterances abandoned before completion. (May or may not be followed by a reformulation.)

    i. He thought that **he have a ...a…he has a** tape decisions.

    ii. He supposed that inside of the suitcase **it..it..** had a lot of money.

Appendix C: Selected original transcript and pruned narrative

Female, EngLow, WR group: 75 seconds long
Uhm After dat *eh* john feel petrified [PAUSE] because[PAUSE] *uhm* and he didn't know what to do. the police come to the[pause] jon and askED [pause] what about your suitcase? PAUSE] you stole that *suitcase* [] john said PAUSE]nonono [pause] *you* are [pause] *ah* confused. Not is, it is not my suitcase is[PAUSE=1.5] that's the stoler[PAUSE] and the police catch the [PAUSE=1.75] the stoler and after that john explained all [pause] the[pause] all the process that happened he say[pause=.39] that he only[PAUSE] he only want to help[PAUSE] *ahh* at the beginning[pause] but he dont know what's the[pause] problem [pause] that the bank was stole[PAUSE] *ehh* [pause] after that[pause] *uhm*[*pause*] he [pause] estole[PAUSE] he killed a beard with *ahh* one stun because it was recognized[pause] as a, [PAUSE] a people that help anothers [pause] and is a hero now

Selected pruned narrative
Female, EngLow, WR group: 75 seconds long: 123 words

After that John feel petrified because

and he didn't know what to do.

The police come to the John and asked, 'what about your suitcase?'

'You stole that suitcase.'

John said no, no, no you are confused.

It is not my suitcase it's…

that's the stoler!

And the police catch the stoler

and after that john explained all the process that happened.

He say that he only want to help at the beginning,

but he don't know what's the problem that the bank was stole.

After that he stole he killed a beard with one stun

because it was recognized as a people that help anothers

and is a hero now.

REFERENCES

REFERENCES

Anderson, J. R. (2009). *Cognitive psychology and its implications*. Worth Publishers.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis. *Review of Educational Research*, *74*(1), 29–58. https://doi.org/10.3102/00346543074001029

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*(2), 390. https://doi.org/10.2307/3587016

Beebe, L. (1983). Risk-taking and the language learner. In H.W. Seliger & M.H. Long (Eds.). *Classroom oriented research* (pp. 39-65). Rowley, MA: Newbury House.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology*, *29*(5), 591-620. https://doi.org/10.1080/00207599408248175

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Chau, H. T. (2014). The effects of planning with writing on the fluency, complexity, and accuracy of L2 oral narratives. (Doctoral dissertation). Retrieved from ProQuest dissertations and Theses database.

Crookes, G. (1989). Planning and interlanguage variation. Studies in Second Language Acquisition, 11, 367-383.

Cumming, Alister & Kantor, Robert & Baba, Kyoko & Erdosy, Usman & Eouanzoui, Keanre & James, Mark. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. Assessing Writing. 10. 5-43. 10.1016/j.asw.2005.02.001.

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. Written Communication, 7, 482-511.

DeBot, K., Chan, H., Lowie, W., Plat, R., & Verspoor, M. (2012). A dynamic perspective on language processing and development. *Dutch Journal of Applied Linguistics*, *1*(2), 188-218. https://doi.org/10.1075/dujal.1.2.03deb

Elder, C., & Iwashita, N. (2005). 8. Planning for test performance. *Language Learning & Language Teaching*, 219-238. https://doi.org/10.1075/lllt.11.14eld

Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, *1*(1). https://doi.org/10.5070/l2.v1i1.9054

Ellis, R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. *Language Learning & Language Teaching*, 167-192. https://doi.org/10.1075/lllt.11.11ell

Emig, J. (1977). Writing as a Mode of Learning. *College Composition and Communication, 28*(2), 122-128. doi:10.2307/356095

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323. https://doi.org/10.1017/s0272263100015047

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, *3*(3), 215-247. https://doi.org/10.1177/136216889900300303

Garcia, P., & Skehan, P. (1999). A cognitive approach to language learning. *TESOL Quarterly*, *33*(4), 769. https://doi.org/10.2307/3587891

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, *11*(4), 329–350. https://doi.org/10.1016/s1060-3743(02)000917

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Hunt, K. W. (1965). Grammatical structures written at three grade levels (Research Report No. 3). Urbana, IL: National Council of Teachers of English.

Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, *37*, 13-38. https://doi.org/10.1016/j.jslw.2017.06.001

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854. https://doi.org/10.1111/lang.12084

Kahng, J. (2017). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569-591. https://doi.org/10.1017/s0142716417000534

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate L2 proficiency. *Language Learning & Language Teaching*, 143-164. https://doi.org/10.1075/lllt.11.09kaw

Kormos, J. (2014). Speech production and second language acquisition. https://doi.org/10.4324/9780203763964

Larsen-Freeman, D. 1983. 'Assessing global second language proficiency' in H. Seliger and M. Long (eds): Classroom Oriented Research in Second Language Acquisition. Rowley, MA: Newbury House Publishers, Inc.

Larsen-Freeman, D. (2006). 6. Functional grammar: On the value and limitations of dependability, inference, and generalizability'. *Language Learning & Language Teaching*, 115-133. https://doi.org/10.1075/lllt.12.08lar

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. https://doi.org/10.1093/applin/16.3.307

Meara, P. (1993). Similar Lexical forms in Interlanguage. Batia Laufer-Dvorkin. Tübingen: Gunter Narr Verlag, 1991. Pp. x 250. DM 124. *Studies in Second Language Acquisition, 15*(1), 122-123. doi:10.1017/S0272263100011748

Levelt, W. J. M. (1989). *ACL-MIT Press series in natural-language processing. Speaking: From intention to articulation.* The MIT Press.

Zeng, L., Mao, Z., & Jiang, L. (2017). The effect of the continuation task on the acquisition of the Chinese spatial phrase structure by L2 Chinese learners. *Chinese Journal of Applied Linguistics*, *40*(3). https://doi.org/10.1515/cjal-2017-0017

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

Manchón Rosa. (2009). *Writing in foreign language contexts: learning, teaching, and research*. Multilingual Matters.

Manchón Rosa. (2011). *Strategies in second language acquisition: a critical assessmant of theory and research*. De Gruyter.

Mao, Z., & Jiang, L. (2017). Exploring the effects of the continuation task on syntactic complexity in second language writing. English Language Teaching, 10(8), 100. https://doi.org/10.5539/elt.v10n8p100

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392. https://doi.org/10.3758/brm.42.2.381

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19-30.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*(1), 83-108. https://doi.org/10.1017/s0272263198001041

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, *12*(1), 11-37. https://doi.org/10.1177/1362168807084492

Mohammad Javad Ahmadian, & Tavakoli, M. (2010). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35-59. https://doi.org/10.1177/1362168810383329

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578. https://doi.org/10.1093/applin/amp044

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*(04). https://doi.org/10.1017/s027226310707043x

Ochs, E. (1979). Planned and unplanned discourse. *Discourse and Syntax*. https://doi.org/10.1163/9789004368897_004

Ortega, L. (1999). Planning and focus on form in l2 oral performance. *Studies in Second Language Acquisition*, *21*(1), 109-148. https://doi.org/10.1017/s0272263199001047

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Peng, J., Wang, C., & Lu, X. (2018). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language Teaching Research*, 136216881878334. https://doi.org/10.1177/1362168818783341

Polio, C. G. (2001). Second language development in writing: Measures of fluency, accuracy, and complexity. Kate Wolfe-Quintero, Shunji Inagaki, and Hae-young Kim. Honolulu: University of Hawai'i press, 1998. Pp. viii+ 187. *Studies in Second Language Acquisition*, *23*(3), 423-425. https://doi.org/10.1017/s0272263101263050

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10-27. https://doi.org/10.1016/j.jslw.2014.09.003

Rahimpour, M., & Safarie, M. (2011). The effects of on-line and pre-task planning on descriptive writing of Iranian EFL learners. *International Journal of English Linguistics*, *1*(2). https://doi.org/10.5539/ijel.v1n2p274

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368-392. https://doi.org/10.1191/026765801681495877

Scardamalia, M., & Bereiter, C. (1982). Assimilative Processes in Composition Planning. *Educational Psychologist*, *17*(3), 165–171. https://doi.org/10.1080/00461528209529253

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Suzuki – Complexity, Accuracy, & Fluency Measures in Oral Pre-Task Planning: A Synthesis From Suzuki (2017)

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*(3), 287-301. https://doi.org/10.1017/s0272263100009177

Wang, C. (2012). The continuation task: An effective way to facilitate L2 learning. Foreign Language World, 5, 2–7.

Wang, C., & Qi, L. (2013). A study of the continuation task as a proficiency test component. *Foreign Language Teaching and Research,* 45 (5), 707-718.

Wang, C., & Wang, M. (2015). Effect of alignment on L2 written production. *Applied Linguistics*, amt051. https://doi.org/10.1093/applin/amt051

Wendel, J. N. (1997). Planning and second language narrative production (Doctoral dissertation). Retrieved from ProQuest dissertations and Theses database. (ProQuest No. 9813575)

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85-106. https://doi.org/10.1177/026553229701400105

Williams, J. N. (n.d.). Working memory and SLA. *The Routledge Handbook of Second Language Acquisition*. https://doi.org/10.4324/9780203808184.ch26

Ye, W., & Ren, W. (2019). Source use in the story continuation writing task. *Assessing* Writing, 39(1), pp. 39-49. https://doi.org/10.1016/j.asw.2018.12.001

Yuan, F. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1-27. https://doi.org/10.1093/applin/24.1.1