Introduction and Brief Summary of Background Information

This study will attempt to address two issues discussed in second language acquisition (SLA) research. The first is a long-standing view of second language (L2)[1] writing as a separate field of research and pedagogy, thus discouraging a view of writing as a means of reinforcing aural discourse (Belcher & Hirvela, 2008). This view is common among instructors and researchers in L2, as it seems to fit intuitively with the most popular current communicative teaching approaches. In a communicative approach, the goal is for the learners to be able to communicate in a given situation. In an attempt to concentrate on developing learners' communicative competence, a traditional linear way of instruction has been accepted; that is, teachers typically focus on practicing reading or listening before speaking in class, while writing is commonly assigned for practice outside of the classroom (Kern, 2000). Thus, this study will look into the effects L2 writing can have on L2 speaking accuracy.

Although the relationship between speaking and writing in SLA research has been limited, researchers have begun to question the view of L2 writing and L2 speaking through the opposite sides of the divide. As Kern (2000) explains, although writing is typically not as bound to spatial and temporal situations as speaking, writing still requires mental construction in a context of interpretation. Additionally, Hyland (2002) suggests that writing, similar to speaking, is an act of communication that requires appropriate interactional manner. Furthermore, Belcher and Hirvela (2008) argue that the current speaking-writing divide can be unproductive for L2 learners because writing not only provides opportunities and a means for learners to practice their L2, but it also encourages learners' explicit attention and noticing. Belcher and Hirvela (2008) also stress that those who teach L2 communication have long been aware of the value of writing in organizing

---

[1] Although the current study is conducted with foreign language learners, the term second language is chosen for convenience due to the prevailing usage of this term in SLA research.

someone's thoughts and ideas before aurally presenting them. Unfortunately, all of these facts have not drawn much attention from L2 researchers, teachers, curriculum designers, or others who may be interested.

Calling for change, Harklau (2002) states that writing should not be marginalized in classroom-based studies and provides an example from her longitudinal study to reinforce this point of view. Following the assumption that "processes of first and second language acquisition can be best traced through careful analysis of classroom talk," Harklau (2002) spent almost a year in L2 classrooms observing L2 learners' interactions with teachers and peers (p. 330). The author was surprised to discover that the learners often interacted with their teacher only once or twice during an entire school day. Their interaction with their native-speaking peers was rarely higher. Instead, the students spent most of their time learning how to read and write. However, the author observed that students had acquired new vocabulary, language forms, and academic content. She concluded that the students acquired the language through print sources of input and practicing written output. Moreover, the learners confirmed that they preferred to work with written sources of input than with spoken input. They explained that written sources were reviewable unlike teacher and peer talk. Thus, it is possible to conclude that the ability to review their own writing also contributed to the process of L2 acquisition. This study makes a significant contribution to the stance that L2 writing can facilitate communicative skills.

Inspired by her study, Harklau (2002) encouraged other researchers to address the current speaking-writing divide more in their future studies. She further asked them "to understand and articulate why – theoretically and methodologically – applied linguists seem much more likely to ask *how students learn to write* in a second language than to ask *how students learn a second language through writing"* (p. 332, emphasis in original). It has become apparent that an act of

writing can influence general L2 proficiency in ways that have not typically been acknowledged. In the pilot study discussed below, the evidence demonstrated that written text reconstruction in L2 was more accurate than in the spoken mode, thus confirming that practicing L2 writing may positively affect speech accuracy. This study will expand to further investigate possible application of L2 writing as a scaffold on L2 speaking. Additionally, the proposed study will offer suggestions for incorporating written texts in an L2 classroom with the explicit purpose of improving aural communicative skills.

The second issue is the prevalent assumption in the field that high individual working memory capacity leads to a better performance in L2 production due to the learners' ability to sustain attention longer, process information faster, and retain knowledge easier (Juffs & Hurrington, 2011). Although a handful of studies investigating the relationship between individual working memory abilities and L2 speaking performance has been conducted, it is still unclear what influence working memory has on L2 speaking skills. One of the explanations for the contradictory results is the absence of agreement of what working memory tests should be used to measure individual working memory abilities.

Studies investigating the relationship between L2 writing and individual working memory abilities are scarce (Juffs & Hurrington, 2011; Bergsleithner, 2010). Although working memory in L2 writing is vastly understudied, the current findings demonstrate a more stable correlation between high working memory capacity and L2 writing performance, but mostly within writing task complexity. One of the explanations of this phenomenon is *the complex cognition hypothesis* (Robinson, 2007) that postulates that individual differences in attentional abilities appear to be most relevant when the cognitive demands of tasks are enhanced.

In the pilot study discussed below, a short and simple text was used for reconstruction tasks in written and spoken mode. The results did not demonstrate a significant correlation between individual working memory capacity and both spoken and written recall accuracy, thus providing more evidence for *the complex cognition hypothesis* proposed by Robinson (2007). According to the cognition hypothesis, individual differences in attentional abilities appear to be most relevant when the cognitive demands of tasks are enhanced. In other words, more complex tasks require increased attentional resources, implying that a high degree of executive attentional abilities will lead to both superior complexity and accuracy. Thus, a cognitively more demanding task is most likely necessary to find a reliable correlation with individual working memory abilities.

The current study will investigate further the relationship between individual working memory abilities and L2 writing and speaking. More complex texts will be used for text reconstruction tasks to check *the complex cognition hypothesis*. Also, English and Russian digit span tasks will be created to measure individual working memory capacity. Comparing the difference in working memory abilities in both L1 and L2 will shed more light on how individual working memory capacity affects the L2 writing and speaking accuracy.

Pilot Study

Although a handful of studies exploring the differences between written recall (WR) and spoken recall (SR) of information have been conducted, to the best of our knowledge, none described the relationship between both recall modes in foreign language learners. With this being said, the purpose of this exploratory study is to analyze the differences between WR and SR in foreign language learners, to determine whether there is the writing superiority effect in L2 learners, and to examine the influence of an individual's WM on both forms of recall. Intermediate-

Low learners of Russian, who speak English as an L1, participated in this study. The following research questions were addressed:

1) What is the relationship between written and spoken recall of narrative texts written in a foreign language?

2) How does individual working memory (WM) capacity influence both types of recall?

It was hypothesized that WR would demonstrate an advantage over SR in terms of accuracy, and that the individual's WM capacity would not affect this advantage. However, it was also hypothesized that high WM capacity would correlate with the accuracy of produced texts in both output modalities due to the benefits of a high WM span. It was assumed that the high degree of executive attentional ability would provide a means for increased accuracy for both encoding and retrieving the information from episodic memory without any additional time needed for editing the produced text.

Since the purpose of this study is to compare WR and SR accuracy but not to explore the differences in cognitive loads required by WR and SR, a time limit was not applied in the recall phase of the experiment to allow opportunities for participants to recall the text as accurately as possible. However, as the results indicated, the actual recall time did not exceed five minutes for both types of recall.

Methodology

Participants

Intermediate-Low learners of Russian (n=22) from two different programs in central USA (n=13; n=9) ranging in age from eighteen to twenty-five participated in the study ($M = 20.64$; $SD = 1.73$). While the participants were not tested for L2 speaking and writing proficiency, it was assumed that they were of similar levels due to being enrolled in classes together. In both groups,

the curricula followed strict *ACTFL* guidelines (2017)[2] for proficiency levels and placement in Russian courses. The same textbooks, '*Golosa*' (Robin, Evans-Romaine, & Shatalina, 2012, books one and two, edition five) were used in the classes for both groups since their first enrollment in Russian courses. At the time of the study, both groups were finishing the first chapter in the second part of '*Golosa*' textbook. The participants received a $20-dollar incentive for their participation. All participants spoke English as their L1.

Materials

A short text (9 simple sentences) was used for the recall task. The text (Appendix 1) was chosen in accordance with the students' level of proficiency, and it included the vocabulary from chapter one of a textbook '*Golosa*' (Robin, Evans-Romaine, & Shatalina, 2012) that was used in both courses of Russian. Using this text ensured that students were familiar with vocabulary and grammatical structures. In order to avoid measuring Russian proficiency instead of the students' ability to recall the text, the text did not include any complex sentences.

The text was typed into a Word document, double-spaced in Times New Roman 12-point font. The measurement of produced written and spoken texts was based on the study of Wolfe-Quintero, Inagaki, and Kim (1998), as adopted by Foster, Tonkyn, and Wigglesworth (2000). Additionally, the measurement of accuracy in both forms of recall was developed for this study (see the Data Analysis and Coding section).

The participants were randomly assigned into two groups based on the output modalities with an equal number of students in each group (n=11): reading the text with written text recall and reading the text with spoken text recall.

---

[2] For more information, see http://www.ACTFL.org/

The English version of the online OSPAN task according to Turner and Engle (1989) was used for measuring individual WMC. In this test, participants had to solve simple mathematical problems while remembering a sequence of letters at the same time. Reading, writing, and speaking span tests were not chosen for the study so as to avoid measuring tasks that are closely related to the experimental tasks used in the study. In other words, in order to avoid negative interference of writing, speaking, or reading on WM results, the OSPAN task was selected. Participants took the test prior to the experiment.

Procedure

The participants were tested individually by the first author. All instructions were read aloud by the experimenter in English and the text version of the instructions was provided to the participants for review. The responses of recall were handwritten (reading-writing mode) or recorded for later transcription by the researcher on an iPhone X (reading-speaking mode).

After signing an informed consent form, each participant took the on-line OSPAN memory test individually. After completing the test, each participant met with the researcher via Zoom. On-line meetings for conducting the experiment were chosen due to unprecedented situation caused by COVID-19. Following safety guidelines, a decision was made not to conduct the experiment in the laboratory to avoid possible exposure of students and the researcher to the virus. During the on-line meetings two scenarios took place. In the first, the participants were told that they would silently read a paragraph during a two-minute time period and would be asked to reconstruct the text on a piece of paper to the best of their abilities. The participants were asked to take a photo of their recalled text and immediately e-mail it to the researcher. The researcher typed the responses in a Microsoft Word document for further analysis. In the second scenario, the participants were told that they would silently read a paragraph during a two-minute time period and would be asked

to reconstruct the text aurally to the best of their abilities over Zoom. The participants were also informed that their response would be recorded on an iPhone by the researcher. All responses were transcribed by the researcher for further analysis.

After completing the presentation phase, the participants immediately proceeded to the second part of the experiment. During this stage, half of the participants were asked to reconstruct the text in writing. The other half were asked to reconstruct the text aurally for a recording. In order to analyze the accuracy of immediate WR and SR, there was not a time limit for either type of recall. However, as previously mentioned, none of the participants spent more than five minutes on the second task.

Data Analysis and Coding

Two trained judges, native speakers of Russian, coded the answers for accuracy and length of the narrations. They discussed all participants' responses until complete agreement on the analysis of the responses (both written and spoken) was reached; in all instances, the consensus was reached. To answer the first research questions, exploring the relationship between written and spoken recall of narrative texts written in a foreign language, the overall accuracy of recall was computed by calculating the number of correct and distorted propositions (T-unit[3] for WR (Wolfe-Quintero et al., 1998) and AC-unit[4] for SR (Foster et al., 2000)) and compared using an independent *t*-test. Any proposition that did not change the meaning of the original proposition

---

[3] T-unit can be defined as 'one main clause plus whatever subordinate clauses attached to it' (Hunt, 1965, p. 20) or 'the shortest unit into which a piece of discourse can be cut without leaving any sentence fragments as residue (Hunt, 1970, p. 189). Examples: When you make a milkshake, you mix it in a blender; He goes to the bookmaker and gets some money (Foster et al., 2000, p. 362).

[4] 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 et al., 2000, p. 365). Examples: That's right; and…er.. they told er… there there was not food crisis (Foster et al., 2000, p. 367).

from the text was counted as a correct proposition. If a proposition provided distorted information compared to the original text, the proposition was counted as a distorted proposition.

Additionally, to measure the syntactic accuracy, the number of reformulations per unit was counted for each produced written and spoken text and also compared using an independent *t*-test. Reformulations were defined as any type of text changes in a sentence produced by the participants that did not change the overall meaning of the sentence. Reformulations included in distorted propositions were discarded from the data. Grammatical accuracy of written and spoken texts and spelling and pronunciation mistakes were not analyzed due to the main objective of checking the learners' memory in L2, but not their L2 proficiency level.

As both written and spoken texts were analyzed, three types of reformulations used by the participants were found: generalization, omission, and substitution. Generalization was used by the participants to replace a certain phrase or a part of the sentence with a possible shorter synonym. For example, instead of "в июне и июле" (in June and July), a generalized phrase "летом" (in summer) was used, or "типичная температура" (typical temperature) could be substituted with "там обычно" (it is usually). Omissions were used by many participants when it was possible to omit a certain word or phrase without changing the meaning in context. For example, "типичная температура в июне и июле – двадцать пять градусов тепла" (the typical temperature in June and July is twenty-five degrees Centigrade) was changed to "типичная температура – *двадцать пять градусов тепла*" (*the typical temperature is twenty-five degrees Centigrade*), "*Летом в Самаре жарко*" (*it is hot in Samara in the summer*) was replaced with "*лето жаркое*" (*summer is hot*). In the case of substitution, possible synonyms or interchangeable words or phrases were used by the participants, e.g., "*Летом в Самаре жарко*" (*it is hot in Samara in the summer*) was replaced with "*Летом там жарко*" (*it is hot there in the summer*).

To answer the second research question, investigating how individual WM capacity influences both types of recall, the relationship between individual WM abilities (an independent variable) and syntactic accuracy (dependent variable), measured by the number of correct propositions, distorted propositions, and reformulations, was evaluated using the *Pearson* correlation test.

Three responses (2SR, 1WR) were discarded from the data due to the fact that zero propositions were recalled. The speaking or writing abilities of the three participants were not sufficient for their recalling cohesive ideas. WM score for the participants was above average (see Table 2), which indicated that the main difficulty with the recall was caused by insufficient knowledge of Russian.

Results

Descriptive statistics of correctly recalled propositions, distorted propositions, reformulations, and the on-line OSPAN test results for both WR and SRs are provided in Figures 2-3 and Table 1. An analysis of mean scores for both types of recall suggests that the WR provided more accurate responses than the SR; i.e., a larger number of correctly recalled propositions ($M = 5.7$, $SD = 1.33$) and a smaller number of distorted propositions ($M = .8$, $SD = .78$) were used in the WR of text as compared to the spoken type of recall ($M = 5.2$, $SD = 2.16$; $M = 2$, $SD = 1.73$). As can be seen in Figure 2, four participants provided zero distorted propositions in written mode of recall, thus directing the value of the standard deviation towards the value of the mean.

Additionally, a smaller number of reformulations ($M = .65$, $SD = .19$) was used by the participants in the written mode of recall as compared to spoken mode of recall ($M = .95$, $SD = .28$).
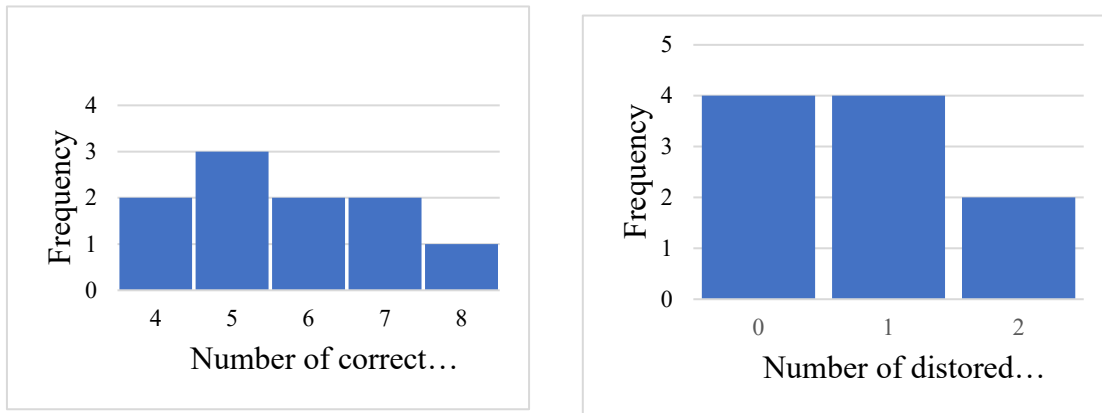

Fig. 1. Frequency of correctly recalled propositions and distorted propositions for the written type of recall
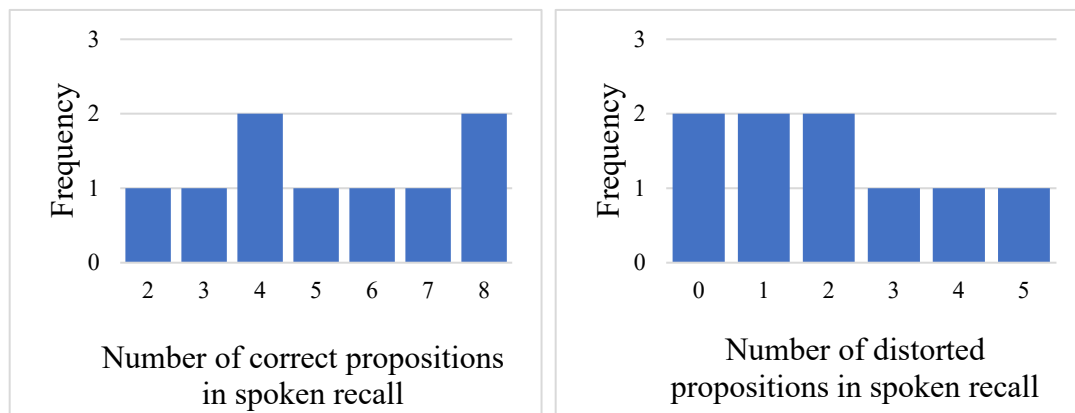

Fig. 2. Frequency of correctly recalled propositions and distorted propositions for the spoken type of recall

Table 1.

Descriptive statistics by number of correct and distorted propositions, number of reformulations, and the WM test total score

	Wrting (*n*=10)		Speaking (*n*=9)	
	M	*SD*	*M*	*SD*
Correct propostions	5.7	1.33	5.2	2.16

Distorted propositions	.8	.78	2	1.73
Reformulations (per unit)	.65	.19	.95	.28
WM score	62	5.8	56	9.7

A two-tailed independent *t*-test conducted for all four variables between two groups (writing and speaking) did not demonstrate a significant difference for the number of correctly recalled propositions ($t(17) = .58$; $p = .56$) and for the WM score ($t(17) = 1.5$; $p = .15$). However, a two-tailed independent *t*-test comparing the number of distorted propositions recalled in written and spoken modes demonstrated a significant difference between WR and SR ($t(17) = -1.9$; $p < 0.05$), indicating more distorted propositions in SR. Moreover, according to an independent *t*-test, the written mode provided significantly fewer reformulations present than found in the spoken mode ($t(13) = -2.75$; $p < 0.05$).

To address the second research question, determining possible correlations between WM individual capacity and accuracy of both types of recall, a *Pearson* correlation test was conducted for individual WMC and three other variables: correct propositions, distorted propositions, and reformulations. The results are provided in Table 2. According to the obtained results, no significant correlation was found in any of the three dependent variables.

Table 2.

Correlation between WM and correctly recalled propositions, distorted propositions, and reformulations for both groups.

	WM_written recall (n=10)	WM_spoken recall (n=9)
	c	*c*
Correct propostions	-.39	-.14
Distorted propositions	-.49	-.27
Reformulations	.55	.33

Note: **Correlation is significant at the 0.05 level (2-tailed)

Discussion

To briefly summarize the results for the first research question, a comparison of WR and SR confirmed the initial hypothesis that WR would produce more accurate responses than SR, and that the differences in individual WMC would not affect the advantage of WR over SR. To be precise, the written responses produced more correct propositions, fewer distorted propositions, and fewer reformulations. Moreover, a significant difference was demonstrated in the number of distorted propositions and reformulations in both types of recall.

Considering the fewer number of distorted propositions and reformulations produced in the written mode of recall and the significant difference in the number of distorted propositions and reformulations produced in both types of recall, the results of this study are consistent with the collective findings from studies with monolinguals, indicating that WR provides more accurate retrieval of information than SR. Writing superiority over speaking in terms of accuracy can be explained by the fact that, unlike speaking, L2 writing activates additional orthographic representations in memory. Activation of additional orthographic representations in L2 writing requires enhanced cognitive resources for both the planning stage and the production stage, thus, leading to increased attention to produced output. Increased attention to output, in turn, leads to better accuracy of produced information. These findings can become a first step for future second language acquisition (SLA) studies analyzing a possibility of scaffolding L2 speaking accuracy by putting more emphasis on L2 writing. This type of research can provide new ideas for improving methods of L2 teaching in the field of instructed SLA.

To summarize the results for the second research question, the WM score did not correlate with the number of correctly recalled propositions, distorted propositions, or reformulations for both types of recall, contrary to the hypothesis. The following explanations of why the higher WM

capacity did not correlate with accuracy are suggested. First, the short length of the text used in the study caused a restriction of range in accuracy of recall. In other words, the participants' responses tended to be highly accurate due to the limited range of recall, which precluded detecting reliable correlations. Perhaps a longer text or a delay in recall is needed to produce sufficient variance in recall scores to show a reliable correlation with the WM score.

Moreover, these results are consistent with the cognition hypothesis proposed by Robinson (2007). According to the cognition hypothesis, individual differences in attentional abilities appear to be most relevant when the cognitive demands of tasks are enhanced. In other words, more complex tasks require increased attentional resources, implying that a high degree of executive attentional abilities will lead to both superior complexity and accuracy. Thus, a cognitively more demanding task is most likely necessary to find a reliable correlation with individual WM abilities.

Limitations

Though WR and SR were tested, the number of the participants was too small to accurately measure both reading and listening input modalities. Thus, the relationship between WR and SR for different input modalities (reading and listening) is still unclear. Moreover, a longer or more complicated written text is recommended in order to receive a more reliable correlation between individual WM abilities and accuracy of both types of recall. Furthermore, studying learners with different levels of proficiency can shed more light on the relationship between WR and SR in L2.

Another limitation of the study relates to the inability to conduct it in a more-controlled laboratory setting. As previously mentioned, due to the unprecedented situation with the spread of COVID-19 followed by the specific safety regulations, a decision was made to conduct the study in an on-line format using Zoom Software. It is hard to predict if the on-line mode affected the results, but future studies could be done in a laboratory or classroom to ensure equal conditions.

In the meantime, the present study demonstrated how research can still be conducted in the most challenging and uncertain times, as COVID-19 global pandemic.

Future Research Directions

The obtained results provide several directions that can be considered for future research. First, the number of the participants in the present study was too small to test the differences in accuracy between WR and SR for two input modalities (reading and listening). Considering the modality congruency effect found in monolinguals (Rummer et al., 2009) and the absence of research analyzing this effect in L2 learners, a study with a 2 x 2 group design (reading input x written output, reading input x spoken output, listening input x written output, listening input x spoken output) should be considered. Second, it is still unclear how exactly WM affects writing and speaking in L2 learners. Studies with more complicated or longer texts can shed additional light as to whether high WM abilities improve L2 learners' accuracy in cognitively more demanding tasks, as suggested by the cognition hypothesis (Robinson, 2007). Third, more studies with learners of different levels of proficiency can provide more insights into the relationship between WR and SR in L2. Finally, considering the overall benefits of writing over speaking in L2 accuracy, future studies can analyze a possibility of scaffolding L2 speaking accuracy by practicing L2 writing. This type of research can provide new ideas for improving methods of L2 teaching in the field of instructed second language acquisition (SLA).

Conclusion

The purpose of this study was to analyze the relationship between WR and SR of episodic memories, and the influence of individual WMC on both types of recall in L2 learners. The obtained results are consistent with results obtained from research on monolingual learners (mostly English). The written mode of recall provided more accurate results than the spoken mode. The

WM score did not correlate, however, with accuracy for both types of recall. This can be due to the low cognitive demands the chosen text required from the participants. A longer text is suggested for future studies comparing WR and SR in L2 learners. Also, participants with a variety of proficiency levels are suggested for future analysis. Finally, SLA studies analyzing a possibility of scaffolding L2 speaking accuracy by practicing L2 writing are encouraged.

Transition to the Current Study

Considering the fact that the written mode of recall was found more accurate in the pilot study, a new direction in SLA research can be suggested. Due to heavier cognitive load and attention required for L2 writing than for L2 speaking, leading to better accuracy of produced texts, it is possible to assume that practicing L2 writing before speaking can provide positive results for improving L2 speaking accuracy. Also, considering the fact that the WM score did not correlate with accuracy for both types of recall, thus confirming the complex cognition hypothesis, the present study will investigate how the differences in working memory processing L1 and L2 influence accuracy of L2 written and aural speech. Another additional factor that the present study will address is how various levels of proficiency of L2 learners affect the relationship between L2 writing and speaking.

Thus, the present study will address the following research questions:

1) How does practicing L2 writing affect L2 speaking accuracy?

2) What effect does a proficiency level have on the relationship between L2 writing and L2 speaking?

3) How does individual working memory capacity in both L1 and L2 affect the relationship between L2 writing and L2 speaking?

Methodology

Participants

Three Novice High and three Intermediate (ACTFL guidelines[5]) learners of Russian participated in the study. The participants were tested for a proficiency level before the study. Only learners with no knowledge of any other Slavic languages participated in the study. The learners participated in the study voluntarily and received a $25-dollar incentive. Participants of each level were randomly split into two groups. One group for each level performed a text reconstruction task following the '*reading-writing-speaking' design*. Other groups performed the same text reconstruction task following the '*reading-speaking-writing'* design.

Materials

Three longer and more complex texts were chosen for each proficiency level. The texts were taken from a Russian textbook '*Shkatulochka*' (Chubarova, 2008) that was not commonly used in the United States (Appendix 2). None of the participants had encountered the texts before. The texts were typed in a Word document, double-spaced in Times New Roman 12-point font and displayed on a computer screen during an individual meeting with the researcher on Zoom. An iPhone XS was used to record the aural responses.

To answer the first research question investigating the relationship between L2 writing and L2 speaking, the participants' responses in both written and aural output modes (independent variables) were analyzed for syntactic and semantic accuracy (dependent measures) (see Coding and Data Analysis section). To answer the second research question investigating how a proficiency level (independent variable) affects the relationship between L2 writing and L2 speaking accuracy (dependent measures), the results for each level of proficiency were compared.

---

[5] For more information, see https://www.actfl.org

To measure individual working memory capacity, on-line *forward and backward digit span tasks* (Olsthoorn, Adriga, & Hulstijn, 2014) have been created in both L1 and L2. Nine digits for each correspondent task have been synthesized by using Speechello speech synthesizer and subsequently edited in PsychoPy (version 3.0) to be exactly 1 second in duration. All digits have been judged by two independent native speakers on their comprehensibility. The auditory forward and backward digit-span tasks in both languages have been created in PsychoPy (version 3.0) software package following the recommendations of Wechsler (1997). The forward and backward digit span task has been chosen for the present study due to the necessity to measure individual working memory capacity in both English and Russian languages. The OSPAN task (Turner & Engle, 1989) used for the pilot study is not offered in the Russian language by the creators.

Participants were instructed that they would hear a series of numbers in English and in Russian and their task was to repeat these numbers out loud for a recording. To familiarize the participants with the digitized numbers, the numbers were played in order from one to nine, while at the same time their visual counterparts were displayed on a computer screen for 1 second. After this stage, participants completed a practice run consisting of three practice trials (two trials of length 2 and one trial of length 3) followed by the experimental part. During this part, participants heard series of digit (first in English, then in Russian) of increasing length and recreated the digits on a computer in the order they are presented for the forward digit span tasks and in reverse order for backward task. The minimum series length was two digits, increasing with one digit every two trials until the maximum length (nine digits for forward series, eight for backward series) was reached, or until participants failed to respond correctly to both trials of a particular length. In total, the working memory test took about 10 minutes to complete.

Procedure

Participants were randomly assigned into two group and tested individually. All instruction was read by the researcher in English and the text version was provided to the participants for review on a computer screen.

Participants completed a consent form, a pre-study questionnaire related to the participants' socio-linguistic background (Appendix 3), and took the forward and backward span tasks on-line at their convenience. Next, they met individually with the researcher on Zoom and two possible scenarios took place. In the first, participants were asked to silently read the provided text in Russian during a five-minute time-period (an approximate time necessary to read each text carefully twice), then to reconstruct the text to the best of their abilities in writing and then aurally. The participants were not given time limit for each reconstruction phase. However, the approximate time required for both reconstruction tasks was no longer than ten minutes. In the second scenario, participants were asked to silently read the provided text in Russian during a five-minute time-period, then to reconstruct the text to the best of their abilities aurally and then in writing. As in the first scenario, the participants were not given time limit for each reconstruction phase. However, the approximate time required for both reconstruction tasks was no longer than ten minutes. When they were done with the reconstruction phase, the participants in both scenarios were asked to answer the post-test questionnaire (Appendix 3). The aural responses were recorded and then transcribed by the researcher. The written responses were typed in a Word document by the researcher for further analysis.

Data Analysis and Coding

Two trained judges, native speakers of Russian, will code the answers for accuracy and complexity of the narrations. The overall syntactic complexity of narrations was computed by

calculating the mean length of clauses per T-unit for the writing tasks (Wolfe-Quintero et al., 1998) and per AS-unit (Foster et al., 2000) for the speaking tasks ($ML_{per}T/AS$) and compared using an independent t-test. In addition, complexity by subordination was calculated by dividing the number of dependent clauses by the number of AS-units or T-units ($DC_{per}T/AS$). Lexical complexity was measured using Guiraud's index, which was calculated by dividing the number of types (verbs, nouns, adjectives, adverbs) by the square root of the number of tokens (total number of words) ($T/\sqrt{2W}$).

To measure the syntactic accuracy, the following error categories were analyzed: functional errors (prepositions, pronouns, conjunctions, word order), verbal errors (subject-verb agreement, conjugations), and nominal errors (gender and plural number agreement in determiners and adjectives), lexical errors (word choice). The number of errors was divided by the number of AS-units or T-units (F_{err}, V_{err}, N_{err}, L_{err} respectively) and compared using the independent t-test.

The overall semantic accuracy of produced narrations was computed by calculating the number of correct and distorted propositions (T-unit for written responses and AC-unit for aural responses) and comparing them using an independent t-test. Any proposition that did not change the original meaning was counted as a correct proposition. If a proposition provided distorted information compared to the original text, the proposition was counted as a distorted proposition.

To answer the third research question, investigating how individual working memory capacity in both L1 and L2 influences L2 writing and speaking, the relationship between individual working memory abilities (an independent variable) and accuracy (dependent variable) was evaluated using the *Pearson* correlation test.

Results

Descriptive statistics of correct and distorted propositions, dependent clauses, mistakes, syntactic complexity, complexity by subordination, and working memory score for each level of proficiency are presented in Tables 3 and 4.

Analysis of the texts produced by Novice learners suggests that the spoken mode produced longer units than the written mode in both design types. The written mode, however, produced more complex sentences by subordination. Lexical complexity, in turn, does not characterize a certain mode of production. As for the accuracy, in both designs the second delivery mode improved overall semantic accuracy of the texts, i.e., the number of distorted propositions decreased. The grammatical accuracy, however, was significantly improved only when the written mode of text production preceded the spoken mode, as it was hypothesized. Although it is impossible to analyze the correlation between working memory capacity and text accuracy and complexity, it is possible to note that working memory capacity in L2 was not as good as in L1, thus signifying that processing L2 for Novice students requires more cognitive load than L1.

Table 3.

Measures of L2 production across writing and speaking modalities produced by Novice learners

	Writing-speaking design (*n*=2)		Speaking-writing design (*n*=1)	
	$M_{writing}$	$M_{speaking}$	$M_{speaking}$	$M_{writing}$
Complexity				
$ML_{per}T/AS$	5.7	6.14	6.7	6.5
$DC_{per}T/AS$	.38	.36	.42	.3
$T/\sqrt{2W}$	2.96	2.83	2.77	2.36
Accuracy				
F_{err}	1	.5	2	3
V_{err}	.5	1	0	1

N_{err}	2	1	5	2
L_{err}	2	.5	1	2
$Total_{err}$	5.5	3	8	8
Correct propostions	9.5	11	10	15
Distorted propositions	3.5	3	4	1
WM English	12.5	12.5	12	12
WM Russian	10	10	10	10

Table 4.

Measures of L2 production across writing and speaking modalities produced by

Intermediate learners

	Writing-speaking design (*n=1*)		Speaking-writing design (*n=1*)	
	$M_{writing}$	$M_{speaking}$	$M_{speaking}$	$M_{writing}$
Complexity				
$ML_{per}T/AS$	13.75	14.89	7.63	6.5
$DC_{per}T/AS$	1.38	1.33	.25	.5
$T/\sqrt{2W}$	4.79	4.95	3.62	4.16
Accuracy				
F_{err}	3	4	3	2
V_{err}	8	1	4	5
N_{err}	6	2	3	4
L_{err}	3	1	0	2
$Total_{err}$	17	8	10	12
Correct propostions	6	9	7	9
Distorted propositions	2	0	1	1
WM English	17	17	14	14
WM Russian	16	16	10	10

Analysis of the texts produced by Intermediate learners suggests that the spoken mode produced longer units than the writing mode in both design types. The written mode, however, produced more complex sentences by subordination. Unlike Novice learners, Intermediate learners produced more lexically complex texts in writing in both designs. As for the accuracy, similar to

Novice learners, the second delivery mode improved overall semantic accuracy of the texts in both designs, i.e., the number of distorted propositions decreased. The grammatical accuracy, however, was significantly improved only when the written mode of text production preceded the spoken mode, as it was hypothesized. Working memory capacity, like for Novice level, was not as good in L2 as it is in L1, thus signifying that processing L2 for Intermediate students also requires more cognitive load than L1.

Discussion

The purpose of this study was to analyze the relationship between L2 writing and L2 speaking accuracy for Novice and Intermediate learners of Russian, and to explore how individual working memory capacity in both L1 and L2 affects this relationship.

Both writing and speaking as the second delivery mode improved overall semantic accuracy of the texts, i.e., the number of distorted propositions decreased. This means that both speaking and writing can provide enough opportunities for leaners to focus on meaning. The grammatical accuracy, however, was significantly improved only when the written mode of text production preceded the spoken mode, as it was hypothesized. This means that L2 learners who engage in writing tasks incorporate uptake from these tasks into speaking accuracy, but not vice versa. This fact can be explained by the increase in attentional resources required for L2 writing. The increased attention, in turn, provides more opportunities for L2 learners to notice forms.

Although practicing writing before speaking improved speaking accuracy for both levels, the written mode provided lexically more complex texts only for Intermediate learners. This fact can be explained by the broader vocabulary range required for the Intermediate level. Novice students are limited in their vocabulary, which prevents them from producing complex sentences. Interestingly, the speaking mode produced longer sentences, but the written mode produced more

complex texts by subordination. Thus, it is possible to conclude that writing provides more opportunities to construct complex sentences for both levels.

Individual working memory capacity in L2 was lower than in L2 for both levels. Therefore, it is possible to assume that processing L2 requires more cognitive load than L1 for both Novice and Intermediate levels. Unfortunately, the low number of the participants did not allow to analyze the correlation between working memory capacity and grammatical accuracy in L2.

Limitations and Future Research Directions

The number of participants was too small to measure correlation of individual working memory capacity in both L1 and L2 with L2 grammatical accuracy. More participants are required to conduct the correlation analysis. The low number of the participants also did not allow to measure accuracy of written and aural texts presented in two different input modalities (reading and listening). Although two output modalities (aural and written) were used for the study, the texts will be presented in only one input modality (reading). It is possible that presenting narration in aural form (listening) would change the results. It is possible that aural reconstruction of text would be more accurate for aural input modality. Thus, studies analyzing different input and output modalities are still warranted.

If the hypothesis of the study is confirmed with a bigger number of participants, further suggestions for L2 pedagogical approaches will be necessary. Considering the sequence of activities in traditional communicative instruction (reading-listening-speaking), new suggestions of how to incorporate writing in a communicative classroom will be needed.

References:

Belcher, D., & Hirvela, A. (2008). *The oral literature connection: Perspectives on L2 speaking, writing, and other media interactions.* Ann Arbor: University of Michigan Press.

Bergsleithner, J. M. (2010). WMC and L2 writing performance. *Ciências & Cognição, 15*(2), 2-20.

Foster, P., Tonkyn, A. & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics, 21*(3), 354–375.

Hunt, K. (1965). *Grammatical Structures Written at Three Grade Levels. Champaign*, IL: National Council of Teachers of English.

Hunt, K. (1970). *Syntactic Maturity in Schoolchildren and Adults*. Monograph of the Society for Research into Child Development.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing, 12*, 17-29.

Hyland, K. (2002). *Teaching and researching writing*. Harlow, UK: Pearson.

Juffs, A., & Hurrington, M. (2011). Aspects of WM in L2 learning. *Language Teaching, 44*(2), 137-166.

Kern, R. (2000). *Literacy and language teaching*. Oxford: Oxford University Press.

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and nonnative speakers. *International Journal of Bilingualism*, 18, 663-673.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics*, 45, 193–213.

Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wolfe-Quintero, K., Inagaki, S. I., & Kim, H.-Y. (1998). *Second language development in writing:*

Measures of Fluency, Accuracy and Complexity. Honolulu, HI: Second Language Teaching

and Curriculum Center, University of Hawai'i at Manoa.

Appendix 1
Text
Note: Spaces have been removed from between the lines.

Сама́ра – э́то го́род на юго-восто́ке Росси́и. Сама́ра – тёплый го́род для Росси́и. Ле́том в Сама́ре жа́рко. Наприме́р, типи́чная температу́ра в ию́не и ию́ле – два́дцать пять гра́дусов тепла́. Ле́том в э́том го́роде мо́жно купа́ться и загора́ть. Весно́й и о́сенью в Сама́ре тепло́ и ча́сто идёт дождь. Зимо́й там хо́лодно и ча́сто идёт снег. Типи́чная температу́ра зимо́й - двена́дцать гра́дусов моро́за. Зимо́й в Сама́ре мо́жно ката́ться на лы́жах и на конька́х. [Samara is a city in the southwestern part of Russia. Samara is a warm city for Russia. It is hot in Samara in the summer. For example, the typical temperature in June and July is twenty-five degrees Celsius. You can swim and sunbathe in Samara in the summer. It often rains in Samara in the spring and in the fall. It is cold and it often snows in Samara in the winter. The typical temperature in the winter is minus twelve degrees Celsius. You can ski and ice skate in Samara in the winter.]

Appendix 2
Advanced Text
Note: Spaces have been removed from between the lines.
Она́ о́чень краси́вая. Когда́ я иду́ с ней по у́лице, на неё все обраща́ют внима́ние. Же́нщины смо́трят с за́вистью, мужчи́ны – с восхище́нием. Молоды́е лю́ди остана́вливаются, подхо́дят к ней, предлага́ют ей ты́сячи ме́лких услу́г: довезти́ на маши́не до до́ма, донести́ су́мку с проду́ктами, проводи́ть на другу́ю сто́рону у́лицы, почини́ть зо́нтик, телеви́зор, магнитофо́н, пригласи́ть в кино́ и́ли в кафе́, угости́ть моро́женым. В авто́бусе ей все уступа́ют ме́сто. Она́ всем нра́вится, у неё мно́го друзе́й. Колле́ги на рабо́те её уважа́ют, нача́льство це́нит.
Она́ умна́ и тала́нтлива. Всё зна́ет и всё по́мнит. Говори́т на трёх языка́х. Игра́ет на гита́ре, поёт, лю́бит поэ́зию, чу́вствует му́зыку, разбира́ется в жи́вописи. Хо́дит на все вы́ставки, конце́рты и кинофестива́ли.
Она́ о́чень везу́чая. Поднима́ет тру́бку телефо́на, набира́ет но́мер спра́вочной слу́жбы – там никогда́ не за́нято. Подхо́дит на остано́вку – такси́ уже́ ждёт. Если покупа́ет лотере́йный биле́т, обяза́тельно выи́грывает. Е́дет отдыха́ть к мо́рю – ни одного́ дождли́вого дня. Оди́н раз в жи́зни идёт на дискоте́ку и там знако́мится со мной.
Она́ до́брая, забо́тливая и внима́тельная. Я то́лько поду́маю, а она́ уже́ говори́т мне: «Что ты сиди́шь до́ма в таку́ю прекра́сную пого́ду? Пойди́ погуля́й, сходи́ с друзья́ми в рестора́н и́ли на футбо́л».
Она́ приве́тливая, гостеприи́мная и скро́мная. Всегда́ уме́ет поддержа́ть у́мный разгово́р. Пре́жде чем что́-то сде́лать, она́ снача́ла смо́трит на меня́, ду́мает и то́лько пото́м принима́ет реше́ние и де́лает.
Она́ молода́я и стро́йная, жизнера́достная и весёлая, мечта́тельная и практи́чная. С ней никогда́ не ску́чно. Она́ лю́бит то́лько меня́. Вот кака́я де́вушка мне нужна́... Вы случа́йно не зна́ете таку́ю?!

Intermediate Text
Note: Spaces have been removed from between the lines.

Преподава́тель сказа́л студе́нтам-иностра́нцам: «Вам не хвата́ет пра́ктики, Вам ну́жно внима́тельно слу́шать, что говоря́т друг дру́гу ру́сские на у́лице, в метро́, в магази́не. По доро́ге в институ́т не повторя́йте дома́шнее зада́ние, а слу́шайте живу́ю ру́сскую речь! Вот вам зада́ние: по́сле уро́ков вам ну́жно пойти́ на остано́вку авто́буса и там постоя́ть полчаса́ и послу́шать, что говоря́т москвичи́. А за́втра расска́жете мне, что вы услы́шали.
Но по́сле заня́тий Па́трик сказа́л: «Мне на́до в аэропо́рт. Сего́дня прилета́ет мой друг Ро́джер. Он никогда́ ра́ньше не́ был в Москве́. Я обеща́л ему́, что встре́чу его́ в аэропорту́». Агне́сса сказа́ла: «А мне на́до в посо́льство. Вчера́ мне позвони́ли отту́да и сказа́ли, что я должна́ получи́ть докуме́нты». Пабло вспо́мнил, что ему́ на́до к врачу́: «У меня́ боли́т нога́. Мне нельзя́ до́лго стоя́ть. Ходи́ть мне то́же нельзя́. Мне мо́жно то́лько е́здить на такси́.» Джулье́тта спеши́ла в парикма́херскую, потому́ что за́втра до́лжен прие́хать в Москву́ её жени́х. А Нико́ль никому́ ничего́ не сказа́ла и куда́-то исче́зла сра́зу по́сле уро́ка. И то́лько Ге́нрих никуда́ не спеши́л. Он пошёл на остано́вку авто́буса. Что же он там услы́шал?

Novice – High Text
Note: Spaces have been removed from between the lines.

Э́то Кири́лл. Его́ жена́ - полице́йский. Она́ мно́го рабо́тает. Она́ ча́сто рабо́тает но́чью. Ве́чером она́ идёт на рабо́ту, а у́тром идёт домо́й спать. Днём она́ спит, ве́чером за́втракает, но́чью опя́ть идёт на рабо́ту. Э́то не то́лько тру́дная, но и о́чень опа́сная рабо́та. Все хулига́ны и банди́ты зна́ют мою́ жену́. Она́ о́чень хоро́ший полице́йский, поэ́тому они́ её не лю́бят.
Кири́лл адвока́т. Он встаём у́тром, за́втракает, идёт на рабо́ту. Иногда́ он встреча́ет на рабо́те свою́ жену́. Э́то быва́ет о́чень ре́дко.
Ко́нчено, все хулига́ны и банди́ты зна́ют его́. Он о́чень хоро́ший адвока́т, поэ́тому все хулига́ны и банди́ты его́ лю́бят. Он хорошо́ их защища́ет, потому́ что э́то его́ рабо́та. Но Кири́лл их не лю́бит, потому́ что они́ не лю́бят его́ жену́. А Кири́лл её о́чень лю́бит.

Appendix 3
Pre-Study Questionnaire
Note: Spaces have been removed from between questions.

Number (given by researcher): _____
Date: _____
1. What is your age? _____
2. When did you begin studying Russian?
3. How many years have you been studying Russian? Have you been studying it continuously? Please mention any breaks in your study.
4. Have you studied abroad for Russian? For how long, in what program and in what city? How long ago was it?
5. What grades have you received in your Russian courses at [*university*] and elsewhere? Please list each course and the corresponding letter grade.

6. How much time, on average, do you devote to studying Russian every week?
7. What other foreign languages have you studied and for how long?
8. Do you have any other exposure to Russian on a regular basis? (i.e. family, friends, TV, etc.) If yes, please mention how often you interact with these people or things.
9. What aspects of Russian are the most difficult for you?
10. Do you think your writing is better than your speaking in Russian or vice versa?
11. Do you expect any benefits from knowing Russian for your future professional career?

Post-Study Questionnaire
Note: Spaces have been removed from between questions.

Number (given by researcher): _____

Date: _____

1) Had you read this text before you participated in this study?
2) Did you pick up any new words or grammar structures from the text?
3) Did the text include any unfamiliar words or grammar structures?