Available online at:
https://dergipark.org.tr/eltrj/
International Association of Research in Foreign Language Education and Applied Linguistics
ELT Research Journal
2022, 11(2), 155-174
e- ISSN: 2146-9814

Investigating the Relationship Between ELT Students' Verbal Working Memory Capacity, Reasoning Ability, and Foreign Language Proficiency of Productive Skills

Ali İlya [a 1], Burcu Koç [a 2], Kerem Can Alpay [a 3]

[a] *Sakarya University, Sakarya, Türkiye*

Research Article Received: 23/06/2022 Accepted: 20/12/2022

To cite: İlya, A., Koç, B,. & Alpay, K.C. (2022). Investigating the relationship between ELT students' verbal working memory capacity, reasoning ability, and foreign language proficiency of productive skills. *ELT Research Journal*, *11*(2), 155-174.

Abstract

The role of working memory and reasoning ability in foreign language proficiency has already been well-established. As another contribution to the building block of the relevant research, the current study aimed to investigate any possible relationships between English Language Teaching (ELT) students' verbal working memory capacity, grammatical reasoning ability, and their proficiency in productive skills in English. 40 ELT students, 20 of whom were in the first-year while the other 20 were in the fourth year, voluntarily participated in the research. They were asked to take two tests on the website of Cambridge Brain Sciences; the Digit Span Test to measure their verbal working memory capacity and the Grammatical Reasoning Test to measure their reasoning ability. They were also tested on their speaking and writing skills in English through TOEFL-IBT test items. The scores on language tests were compared with the scores on the Digit Span and the Grammatical Reasoning Tests by running multiple regression analysis and a full-factorial ANOVA. Results revealed that while the writing ability had a relationship with the verbal working memory and reasoning ability, the speaking ability could not be predicted by them. Moreover, it was also yielded that the grade level of the students did not have any effects on the L2 productive skills.

Keywords: Verbal working memory; reasoning ability; L2 productive skills; digit span

[1] Corresponding author.
E-mail address: ailya@sakarya.edu.tr
[2] Second Author: burcukoc@sakarya.edu.tr
[3] Second Author: keremcan@sakarya.edu.tr

Introduction

The role of memory in the domain of foreign language learning and use has already been well-established on the basis of the results yielded by the research studies in Second Language Acquisition (SLA) (Acheson & MacDonald, 2009; Baddeley, 2003; Daneman & Merikle, 1996). The association between modalities of language and different types of memory has been analyzed from diverse perspectives. A large number of studies have focused on the dimension of language comprehension in relation to memory while a smaller proportion is on language production. More specifically, memory, either as a unitary system or as a multidimensional model, has been explored in relation to separate layers of both language learning and use; vocabulary (Jefferies, 2006a/b; Schriefers et al., 1990; Walker & Hulme, 1999), phonology (Vousden et al., 2000; Schwartz, 2004; Service, 1992), reading (Daneman & Carpenter, 1980; McCutchen, 1991), listening (Just, 1992), and speaking (Vitevitch, 2002), to list a few. All these studies have somehow revealed a certain interaction between individuals' separate types of memory capacity, reasoning abilities, and success in language learning and use. Still, as is also asserted by Acheson and MacDonald (2009), memory-related studies in the field of language research have predominantly targeted either comprehension or acquisition, and the productive modalities have remained underrated in terms of exploration within that scope. Moreover, the number of studies generated in the Turkish context, especially with people whose main area of specialization is a foreign language teaching, is quite limited. As a response to this need, the current research study attempts to investigate any relationships between proficiency in the second language (L2) productive skills and verbal working memory (VWM) as a specific type, and reasoning skills of adults pursuing a Bachelor of Arts (BA) degree in the field of English Language Teaching (ELT) at a state university in Turkey. To introduce the related concepts and terms and better locate the results yielded here into the body of findings put forward so far, a concise review of the literature is a requisite.

Review of Literature

Memory, as a broadly discussed phenomenon, is initially divided into two essential categories; short-term memory (STM) and long-term memory (LTM) based on the suggestion by Hebb (1949). In another study, it is also empirically confirmed that elements a person is exposed to rapidly go off if there is not any rehearsal (Brown, 1958), and it signals the existence of a separate temporal storage system. The upcoming decade welcomes arguably the most commonly accepted two-dimensional model of memory proposed by Atkinson and Shiffrin (1968). Their model introduces short-term storage as 'an antechamber to the more durable

LTM' (Baddeley, 2003). They do not confine the function of a short-term storage system to a bridge leading to long-term memory but identify it as working memory which operates as a primary asset to many sophisticated activities, reasoning, and comprehension. Later, Baddeley and Hitch (1974) further divide the short-term storage system into three subcomponents that complement each other as separate parts in the process of dealing with complex tasks but are bound with the same unitary system named as working memory. The central executive is defined as the attentional control system that administers the visuospatial input from the visuospatial sketchpad and phonological data from the phonological loop.

Subsequent advancements in research unfold two main deficits in the three-component working memory system (Baddeley & Logie, 1999). The existing model assigns a distinctive separate storage system for visual and verbal input; however, the necessity of another part that blends the two sorts of information and forwards them to the multidimensional representations in LTM arises. Additionally, it is revealed that the overarching quantity of input to be temporarily stored far exceeds the capacity of either the phonological loop or visuospatial sketchpad; that's why another compartment within the STM is required (Baddeley, 2003). Out of these recently emerging realizations, the fourth component of working memory is born; the episodic buffer. Its assumed function is to compound input from two other subsystems with one another and to transmit it as an integrated chunk to the long-term memory. Figure I visually represents the structure of the newly shaped working memory.

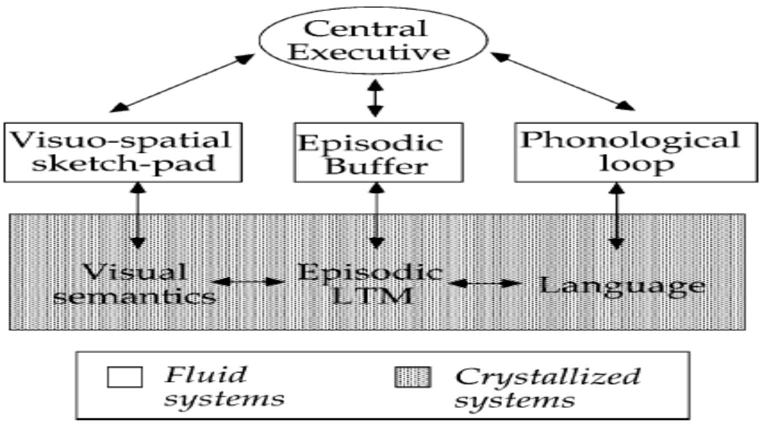

Figure 1. The four-component model of working memory (Baddeley, 2003, p. 203)

As can be seen in Figure 1, language abilities are closely associated with the phonological loop. It functions for 'storing and processing verbal and acoustic information' (Tsai, 2014). It is comprised of two sub-components, as well. The first subcomponent saves information solely for a few seconds, and then the input quickly fades away without being consolidated or reactivated by another item. The second sub-component, the subvocal rehearsal system, stores the information and records any visual stimulus under the condition that it can be named (Baddeley, 2003). The ease and span of retention of the input rely on a few variables as yielded by the studies delineating the relationship between the phonological loop and language performance; phonological and/or semantic similarity of the items, length of the words, recency, and primacy (Acheson & MacDonald, 2009; Baddeley, 2003). Depending on these factors, the items are linked with long-term memory, as displayed in the 'crystallized' area in Figure 1.

In relation to the working memory model, a significant percentage of studies have addressed the VWM from various angles in the field of language studies. It has been specified that VWM possesses a crucial role in the development of reading skills, vocabulary learning, and overall proficiency in a language (Harrington & Sawyer, 1992; Service & Kohonen, 1995). Daneman and Merikle (1996) indicate that, for successful comprehension, newly encountered input is to be incorporated into the previously processed information, and thus it necessitates access to the existing storage, which means reading and listening comprehension skills can justifiably be associated with VWM capacity. Research studies, for a long amount of time, have utilized traditional measures of VWM such as digit span, word span, and letter span tests. The results of the studies investigating the correlation between working memory capacity and scores on standardized tests of comprehension have been disappointing for scholars (Daneman & Carpenter, 1983; Mitchell, 1982; Just & Carpenter, 1980). As discussed by Daneman and Merikle (1996), the lack of correlation between the two variables is attributed to the problems with the traditional measures by Daneman and Carpenter (1980). They claim that the traditional tests measure only the storage capacity; however, the capability of working memory depends more on functionality rather than storage. Thus, with the newly developed test by Daneman and Carpenter (1980), the previous studies are replicated, and the results are in contrast with those of the aforementioned studies (Daneman & Merikle, 1996). The correlation between the score on VSAT-Verbal Scholastic Aptitude Test and the scores on listening and reading comprehension tests is .66 on average (Daneman & Carpenter, 1980). The follow-up studies, conducted through various revised forms of VSAT, support the correlation between the two variables (Gaulin & Campbell, 1994; Leather & Henry, 1994).

Studies dealing with the comprehension modalities significantly outnumber those addressing the relationship between productive skills and VWM capacity (Acheson & MacDonald, 2009). However, it has been observed that errors in speech production are parallel to the errors on the tests of VWM (Ellis, 1980), and hence it suggests that studies which investigate the relationship between someone's VWM capacity and his/her productive performance in a language are worth as much attention as those that examine the connection of VWM capacity and comprehension skills (Acheson & MacDonald, 2009). The growing need for exploring the relationship between productive skills and VWM paves the way for an increase in the number of studies within that category. Weissheimer (2011), in a longitudinal study, addresses the interaction between speaking span test scores and working memory capacity during L2 learners' language development process. Results display that only lower-span people have a statistically significant improvement in their working memory scores, but both high and moderate-span participants' WM scores increase to some extent. In an older study, Daneman (1991) reveals that fluency in speech generation positively correlates with the participants' skills of processing and temporary storage functions of WM.

Even though the relationship between the VWM capacity and L2 production skills has already been investigated (Grabowski, 2007; Kim & Tracy-Ventura, 2011; Mackey et al., 2010, McCutchen, 2000; Révész, 2012), the number of such studies is still quite limited. To illustrate, even though working memory operates relatively differently in speaking and writing, higher working memory capacity seems to improve access to language and information processing in both speaking and writing (Mackey et al., 2010, McCutchen, 2000). To provide further clarification, Mota (2003) has spotted a correlation between enhanced working memory capacity and improved L2 performance in terms of complexity, accuracy, and fluency, but not lexical density. On the other hand, while it is widely accepted that having a large working memory flourishes language performance, there are conflicting findings about the circumstances in which this advantage reveals itself (Cho, 2018). To exemplify, studies conducted by Tavares (2009) and Ahmadian (2012) indicate that a large working memory only offers an advantage in oral output when sufficient time is available for planning.

Moreover, in the Turkish context, it has been quite rarely studied, and given the influence of discrete first languages (L1s) on L2 development, studies specifically addressing the Turkish people might put forth different results. In addition, those who are studying language professionally may be more inclined to score better on VWM and reasoning tests, which might be quite a clear indicator of the relationship between the two variables. For these reasons, the current study is conducted to fill the explained gap.

The following research questions guide this study:

1. Are there any relationships between ELT students' proficiency in L2 productive skills and their VWM and reasoning test scores?

2. Do first-year ELT students differ from fourth graders in terms of the relationship between proficiency in L2 productive skills and their VWM and reasoning test scores?

The research questions are produced on the basis of the null hypotheses as follows:

1. There is no relationship between ELT students' proficiency in L2 productive skills and their VWM and reasoning test scores.

2. First-year ELT students do not differ from fourth-graders in terms of the relationship between proficiency in L2 productive skills and their VWM and reasoning test scores.

Methodology

Research Design

The researchers embraced a post-positivist worldview to find answers to the research questions. Post-positivism features a conjectural view of knowledge, deterministic philosophy, and a reductionistic approach (Creswell, 2014; Philips & Burbules, 2000). In line with this worldview, a quantitative orientation was adopted. A cross-sectional survey design was employed for the reasons of generalizability from a sample to a population, economy, and practicality (Creswell, 2014).

Participants

The population of the study was composed of undergraduate first-year and fourth-year students in the departments of ELT in Turkey. The sample encapsulated 40 undergraduate ELT students who were studying at a large-scale, state university in Turkey. Participants were selected among the 1st and 4th year students at equal number 20 on the basis of the principles of convenience sampling as a type of nonprobability sampling method. Eighteen of the participants were male, while twenty-two were female. They were formerly informed about the general outline of the research study and asked to sign an informed consent form to admit their voluntary participation in a written form. The language proficiency level of the students was either equal to or above B2 according to the Common European Framework of Reference for Languages (CEFR), depending on the scores they achieved to pass or be exempted from the compulsory preparatory education. In order to be deemed successful here, all the newly enrolled students were required to get 80 out of 100 on a four-section test, in which each section measured their ability in a separate component of English. The test was at the level of TOEFL IBT, and according to the regulations approved by the Senate of the University, an 80 was equal to B2 on CEFR and 96 out of 120 on TOEFL IBT. Thus, it could justifiably be claimed that all the participants were at B2 or a higher level according to the levels of CEFR.

Data Collection

In order to measure the participants' proficiency in productive skills in English, the Internet-Based Form of the Test of English as a Foreign Language (TOEFL IBT) was used. Digit Span Test was employed to measure VWM capacity, while Grammatical Reasoning Ability Test was performed to measure participants' reasoning ability.

TOEFL IBT is an international test administered by Educational Testing Service (ETS). It measures the proficiency level of four language skills i.e. reading, writing, listening, and speaking separately. In line with the scope of this study, writing and speaking sections would be introduced here. The writing section of the test is divided into two as independent writing and integrated writing. Likewise, the speaking section also encompasses two distinct parts as independent speaking and integrated speaking. For the current study, independent writing and speaking tests were given. In the independent writing test, test takers are asked to write an essay in 30 minutes based on their personal opinion and experience. In the independent speaking test, the test takers are asked to speak about a personal topic by drawing completely on their personal experience and opinion. Test takers are given 15-30 seconds to prepare their responses and 45-60 seconds to speak. The reliability coefficients provided by the test developers are 0.86 for the speaking section and 0.80 for the writing section.

Digit Span Test and Grammatical Reasoning Test are offered by Cambridge Brain Sciences. Daneman and Merikle (1996) list the Digit Span Test among the traditional measures of VWM. It was designed to measure verbal short-term and working memory. It is presented in two formats as Forward or Reverse Digit Span Test. The test-takers are given a random series of digits and required to repeat them in the same order (forward) or in the reverse order (backwards). Upon the correct trial of the test-taker, a longer sequence is presented, and the same verbal repetition in both directions is asked. The test ends when the test-taker fails to accurately remember the sequence of the digits on three occasions. The test-taker's span is estimated based on the longest sequence of numbers accurately remembered. The test lasts 1-3 minutes based on the test-taker's performance. Grammatical Reasoning Test measures one's ability to reason about relationships among objects. It measures participants' verbal reasoning (Hampshire et al., 2012). The test is the adaptation of the original Grammatical Reasoning Test developed by Baddeley (1968). In the test, a statement appears at the top of the screen to describe the relationship among the shapes or objects beneath, and the participants are asked to indicate if the statement accurately describes the shapes by clicking 'true' or 'false'. Participants have 90 seconds to complete the test. The number of correct answers constitutes the participants' scores on the test.

The detailed description of the procedures for the collection of data through these four instruments is as follows:

- TOEFL IBT-Independent Writing Test: All the students were asked to compose a well-developed essay in a sample test offered by ETS. The topic was the same for all the participants. The responses were graded according to the set of criteria defined in the rubric of ETS. Each student got a score from 1 to 5 according to the range in the rubric and all the scores were converted into marks out of 100. Before grading the papers, the researchers piloted grading a few papers that were out of the scope of the research study. Following an agreement on the standards defined in the rubric, each student paper was graded by two of the researchers, and in the case of an inconsistency that is greater than 10% the third researcher was consulted.
- TOEFL IBT-Academic Speaking: The students were asked to talk about three different topics, one after another. They were given 15 seconds to prepare their response and 45 seconds to speak as it is in the actual TOEFL IBT test. Responses were graded in accordance with the criteria set in the rubric prepared by ETS, and the scores were converted into marks out of 100. The tests were the samples provided by ETS. Prior to the actual tests, the researchers piloted the testing process and agreed on the standards in the rubric. Two independent raters graded a performance. A third researcher was consulted when the gap between the two grades was higher than 10%.

- Digit Span Test: All the participants were kindly requested to create an account on the website of Cambridge Brain Sciences. Then, in a language laboratory, they were allotted an individual computer to take the test. They were introduced to the test procedures and any questions were answered. They were initially permitted to take the test once to get themselves familiarized with the content of the test, and then they repeated the test for the purpose of the research. They got a screenshot of the score they had and sent it to the researchers via e-mail.

- Grammatical Reasoning Test: The same procedures were followed with the Digit Span Test. The participants took the Grammatical Reasoning Test subsequent to a short break following the completion of the Digit Span Test.

Data Analysis

In line with the nature of the quantitative research methodology, the researchers conducted a multiple regression analysis so as to reveal if any of the independent variables accurately predicted the writing and/or speaking performance of the participants. Additionally, a full factorial 2x2 ANOVA test was run to see the relationship among the variables. IBM SPSS was used for data analysis. The assumptions of the multiple regression were tested. The linearity of dependent variables, homoscedasticity of data, independence of independent variables, and normality of distribution were all maintained.

Results

The arithmetic mean of the participants' writing scores was found to be 62.75, and for the speaking test, it was 62. The writing scores range between 30 and 90 while the speaking scores were between 35 and 90. The lowest score on the Grammatical Reasoning Test was 7 and the highest was 20. For the Digit Span Test, the scores ranged between 5 and 14.

Table 1. *Descriptive Statistics for the Test Scores*

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
GR_Test_Score	40	7	20	11,23	2,547	,911	,374	2,291	,733
Digit_Span_Scr	40	5	14	8,35	2,070	,982	,374	,731	,733
Writing_Score	40	30	90	62,75	13,726	-,033	,374	-,051	,733
Speaking_Score	40	35	90	62,00	14,754	-,168	,374	-,832	,733

Valid N (listwise)	40

In order to test if any of the independent variables; the score on either the Digit Span Test, the score on the Grammatical Reasoning Test, or the grade level of the participants significantly predicts the score on the writing test, a multiple regression analysis was conducted. As presented in Table 2, corr(WS, DST)=.43, which indicates that the writing score significantly correlates with the score on the Digit Span Test at the value of p=0.003. Additionally, corr(WS, GRT)=.341, which also reveals a strong correlation between the writing score and the Grammatical Reasoning Test score with the p=0.016. However, it seems that grade level does not significantly correlate with the writing scores of the students with the p<0.05.

Table 2. *Correlations Between the Writing Score and Other Variables*

		Writing_Score	GR_Test_Score	Digit_Span_Scr	Grade
Pearson Correlation	Writing_Score	1,000	,341	,430	-,148
	GR_Test_Score	,341	1,000	,150	,109
	Digit_Span_Scr	,430	,150	1,000	,147
	Grade	-,148	,109	,147	1,000
Sig. (1-tailed)	Writing_Score	.	,016	,003	,182
	GR_Test_Score	,016	.	,178	,251
	Digit_Span_Scr	,003	,178	.	,183
	Grade	,182	,251	,183	.
N	Writing_Score	40	40	40	40
	GR_Test_Score	40	40	40	40
	Digit_Span_Scr	40	40	40	40
	Grade	40	40	40	40

The regression model that is comprised of the digit span and grammatical reasoning test scores along with the grade could accurately predict the writing scores of the participants with the p=0.03 as is displayed in Table 3. The coefficients also show that the writing score could be predicted by the digit span test score at the p=0.005, by the score on the grammatical reasoning test at the *p*=0.036. However, grade level does not predict the writing scores of the participants with the *p*<0.05.

Table 3. *Regression Model Summary for the Writing Score*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,566*	,320	,264	11,778	,320	5,657	3	36	,003

[a] Predictors: (Constant), Grade, GR_Test_Score, Digit_Span_Scr

Table 4. *Coefficients for the Writing Score*

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	30,924	11,360		2,722	,010	7,885	53,964
GR_Test_Score	1,642	,752	,305	2,184	,036	,117	3,168
Digit_Span_Scr	2,784	,930	,420	2,995	,005	,899	4,670
Grade	-6,574	3,780	-,243	-1,739	,091	-14,241	1,093

[a] Dependent Variable: Writing_Score

The same analysis was also administered for the speaking scores of the participants. The results of the correlational analysis did not reveal any powerful relationships between any of the variables as can be seen in Table 5. The values of correlation are all below 0.30.

Table 5. *Correlations Between the Speaking Score and Other Variables*

		Speaking_Score	GR_Test_Score	Digit_Span_Scr	Grade
Pearson Correlation	Speaking_Score	1,000	,196	,174	,017
	GR_Test_Score	,196	1,000	,150	,109
	Digit_Span_Scr	,174	,150	1,000	,147
	Grade	,017	,109	,147	1,000
Sig. (1-tailed)	Speaking_Score	.	,113	,142	,458
	GR_Test_Score	,113	.	,178	,251

	Digit_Span_Scr	,142	,178	.	,183
	Grade	,458	,251	,183	.
N	Speaking_Score	40	40	40	40
	GR_Test_Score	40	40	40	40
	Digit_Span_Scr	40	40	40	40
	Grade	40	40	40	40

The results of the multiple regression analysis also demonstrate that none of the independent variables significantly predict the speaking performance of the participants. Table 6 displays that the regression model is not good enough to accurately predict the speaking scores of the participants with the $p>0.05$. Likewise, the coefficients presented in Table 7 do not indicate any statistically significant relationships between the variables.

Table 6. *Regression Model for the Speaking Scores of the Participants*

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,245*	,060	-,018	14,887	,060	,769	3	36	,519

[a] Predictors: (Constant), Grade, GR_Test_Score, Digit_Span_Scr

Table 7. *Coefficients for the Speaking Scores*

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	42,640	14,359		2,969	,005	13,518	71,762
GR_Test_Score	1,019	,950	,176	1,072	,291	-,909	2,946
Digit_Span_Scr	1,076	1,175	,151	,916	,366	-1,307	3,460
Grade	–,706	4,778	-,024	-,148	,883	-10,397	8,985

[a] Dependent Variable: Speaking_Score

A full factorial 2x2 ANOVA was run to check any interactional effects between the variables for both writing and speaking test scores. The results for the writing test scores are identified in Table 8. As was revealed, no statistically significant interaction was observed between any of the variables with all the values of $p<0.05$.

Table 8. *ANOVA Results for the Writing Test Scores*

Source	Type II Sum of Squares	df	Mean Square	F	Sig	Partial Eta Squared
Corrected Model	6335,000	34	186,324	,920	,615	,862
Intercept	157502,500	1	157502,500	777,790	,000	,994
GR_Test_Score	1326,503	9	147,389	,728	,681	,567
Digit_Span_Scr	1360,717	7	194,388	,960	,538	,573
Grade	232,692	1	232,692	1,149	,333	,187
GR_Test-Score* Digit_Span_Scr	927,762	5	185,552	,916	,537	,478
GR_Test-Score* Grade	7,500	1	7,500	,037	,855	,007
Digit_Span_Scr* Grade	400,000	2	200,000	,988	,435	,283
GR_Test-Score* Digit_Span_Scr* Grade	,000	0	-			,000
Error	1012,500	5				
Total	164850,000	40	202,500			
Corrected Total	7347,500	39				

[a] Dependent Variable: Writing Score

[b] R Squared= ,862 (Adjusted R Squared= -,075)

The results were not different for the speaking scores either. No interactional effect was detected for any of the variables with all the values of $p<0.05$. Only the interaction between the scores on the digit span and the grammatical reasoning tests is close to be statistically significant with the $p=0.113$. The results are displayed in Table 9.

Table 9. *ANOVA results for the Speaking Scores*

Source	Type II Sum of Squares	df	Mean Square	F	Sig	Partial Eta Squared
Corrected Model	7760,833	34	228,260	1,565	,328	,914
Intercept	153760,000	1	153760,000	1054,354	,000	,995
GR_Test_Score	935,755	9	103,973	,713	,690	,562

Digit_Span_Scr	2845,473	7	406,496	2,787	,138	,796
Grade	376,923	1	376,923	2,585	,169	,341
GR_Test-Score* Digit_Span_Scr	2345,436	5	469,087	3 ,217	,113	,763
GR_Test-Score* Grade	3,333	1	3,333	,023	,886	,005
Digit_Span_Scr* Grade	277,083	2	138,542	,950	,447	,275
GR_Test-Score* Digit_Span_Scr* Grade	,000	0	-			,000
Error	729,167	5				
Total	162250,000	40	145,833			
Corrected Total	8490,000	39				

[a] Dependent Variable: Speaking Score

[b] R Squared= ,914 (Adjusted R Squared= ,330)

Discussion and Conclusion

The analyses conducted so as to discover the answers to the research questions revealed that the digit span test that is intended to measure the VWM capacity successfully predicted the writing ability of the participants. It was in parallel with the results of the research conducted by McCutchen (2000), Moa (2003), Mackey et al. (2010), and Révész (2012). Daneman (1991) and Weissheime (2011) additioanlly revealed a correspondence between productive L2 ability and VWM capacity although they mainly dealt with speaking ability. Besides the digit span test, the grammatical reasoning test could also predict the writing scores of the participants. The same tests did not predict the speaking scores of the participants. The results here suggested that the two tests -the digit span and grammatical reasoning- have a certain amount of relationship in terms of the components they measure. The ANOVA also pointed to the interactional effect of the two tests on the speaking score although it was not statistically significant. The processing differences that were revealed could be explained in terms of control, planning, and monitoring writing allows for the language users. Due to these differences, the amount of linguistic information that can be gathered and shown during the performance of a task is limited. L2 writers are better able to retrieve information stored in their long-term memory than L2 speakers because they have a higher command of the language (Grabowski, 2007).

Secondly, it could be inferred that even though both speaking and writing are categorized under productive skills, they are distinct from each other, as revealed by contradictory relationships between these scores and the other two. In line with this, studies conducted by Tavares (2009) and Ahmadian (2012) suggested that planning time permits working memory to govern cognitive aspects by activating pertinent information and inhibiting irrelevant information. These findings indicated that a large working memory may be more beneficial for writing than for speaking, as writing affords more planning opportunities. However, multiple modes of production embrace more than simply variations in planning; thus, it is vital to do empirical studies on the functionality of working memory in diverse forms of production.

The grade level of the students did not seem to be a substantial predictor of the speaking and writing performances. It might be due to the fact that, in the ELT curriculum, courses generally address methodologies or other dimensions of language teaching rather than the language itself. Thus, no or little improvement might be observed in the productive L2 ability of the participants from the 1st to the 4th year.

In conclusion, the first null hypothesis could not be fully rejected as the speaking score did not have any relationship with either the digit span or the grammatical reasoning test scores. However, it was partially rejected due to the interactional effect of the digit span and grammatical reasoning test scores on the writing performance of the students. The second null hypothesis was admitted since no effect of the grade level was revealed by the analyses. Then, it could be concluded that the VWM capacity and reasoning abilities were partially in relation to the productive L2 abilities while grade level had nothing to do with it.

The partial associations between working memory and language performance were consistent with the weak influence of working memory on total language performance (Juffs & Harrington, 2011). As language performance is influenced by a variety of contextual and performative elements, the impacts of other performance variables such as strategy, context, and familiarity that may outweigh those of working memory should be incorporated into further research (McCutchen, 2000). There is evidence supporting the positive effects of working memory on long-term language development (Kormos & Sáfár, 2008), so it would be beneficial to examine the effects of working memory on performance over an extended period of time or in conjunction with other production-related variables such as attitudes and strategies.

The Research and Publication Ethics Statement

The authors declare that data for the study was collected before 2018. Therefore, Ethics Committee Approval is not an obligation for this study.

The Conflict of Interest Statement

In line with the statement of Committee on Publication Ethics (COPE), I/we hereby declare that I/we had no conflicting interests regarding any parties of this study.

Contributions of authors

All authors have contributed to all stages of the research process including the implementation, data analysis, reporting, editing, and any other.

Acknowledgment

This study was presented at "12th International ELT Research Conference: Analyze the Present, Imagine the Future," May 16-18, 2022, Çanakkale/TÜRKİYE.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50.

Ahmadian, M. J. (2012). The relationship between working memory capacity and L2 oral performance under task-based careful online planning condition. *Tesol Quarterly*, *46*(1), 165-175.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), T*he Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89–195). New York: Academic Press.

Baddeley, A.D. (1968). A three-minute reasoning test based on grammatical transformation. *Psychometric Science, 10*, 341-342

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829-839.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 28–61). New York: Cambridge University Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12–21.

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System*, *72*, 85-98.

Creswell, J. W. (2014). *Qualitative, quantitative and mixed methods approaches*. Singapore: SAGE Publications.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*(6), 445-464.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning. Memory. & Cognition, 9*, 561584.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422-433.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning & Verbal Behavior*, *19*, 624–634.

Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. *Studıes in Wrıtıng, 20*, 165.

Gaulin, C. A., & Campbell, T F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual & Motor Skills, 79*, 55-64.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron, 76*(6), 1225-1237.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(01), 25-38.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Jefferies, E., Frankish, C., & Ralph, M. A. L. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59*(5), 949–964.

Jefferies, E., Frankish, C. R., & Ralph, M. A. L. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54*(1), 81–98.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. Language teaching, 44(2), 137-166.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 4*, 329-354.

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

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*(2), 261-271.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*, 88-111.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language learning*, 60(3), 501-533.

McCutchen, D., Dibble, E., Blount, M. M. (1991). Phonemic effects in reading comprehension and text memory. *Applied Cognitive Psychology, 8*, 597–611.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist*, *35*(1), 13-23.

Mitchell, D. C. (1982). *The process of reading: A cognitive analysis of fluent reading and learning to read.* New York: Wiley.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos: Revista de Língua e Literatura Estrangeiras*, *24*.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. New York: Rowman & Littlefield.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, *62*(1), 93-132.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86-102.

Schwartz, M.F., Wilshire, C.E., Gagnon, D.A., Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology, 21*, 159–186.

Service E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *45*(A), 21–50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquis ition? *Applied Psycholinguistics*, *16*, 155-172.

Tavares, M. D. G. G. (2009). The relationship among pre-task planning, working memory capacity, and L2 speech performance: a pilot study. *Linguagem & Ensino, 12*, 165–194.

Tsai, A. M. (2014). The role of visuo-spatial and verbal working memory in L2 Japanese reading proficiency. *University of Hawai'I Second Langauge Studies Paper 32 (2)*.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 735–747.

Vousden, J.I., Brown, G.D.A., Harley, T.A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101–175.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental psychology: Learning, Memory, and Cognition*, *25*(5), 1256.

Weissheimer, J. (2011). The role of working memory capacity in the development of L2 speech production. *A Journal of English Language, Literatures in English and Cultural Studies, (60*), 075-104.