Working Memory and Note Quantity: Their Relationship with Consecutive Interpreting in Proficient Bilinguals. Implications for Aptitude Tests of Interpreting

Mojtaba Amini | Azizollah Dabaghi | Dariush Nejadansari
m.amini257@yahoo.com | dabaghi@fgn.ui.ac.ir | ansari@fgn.ui.ac.ir
The University of Isfahan

Recibido: 13/01/2020 | Revisado: 25/03/2020 | Aceptado: 30/05/2020

Abstract

The present study examined the association between working memory (WM) and note quantity and their relationship with consecutive interpreting (CI) performance in order to evaluate their predictive efficiency for aptitude tests in CI. To follow the objectives of the study, two WM tests and one CI task were administered to 30 MA translation students. The results indicated a positive and significant relationship between one of the WM measures (Reading Span) and note quantity. Moreover, there was a significant relationship between both measures of WM and CI performance, and also between note quantity and CI performance. Furthermore, the results of the regression analysis indicated that both measures of WM were predictors of CI performance while the variable ‘note quantity’ failed to be a predictor. Based on the results, it was further proved that WM is an efficient component in aptitude tests whereas note quantity did not completely comply with the criteria and was rejected as a reliable criterion and could not be included as a subcomponent in the aptitude tests.

Keywords: Consecutive Interpreting (CI), Working Memory (WM), Note-taking, Note quantity, Aptitude test

Resumen

Memoria de trabajo y cantidad de notas: su relación con la interpretación consecutiva en bilingües competentes. Implicaciones para las pruebas de aptitud de interpretación

El presente estudio examina la asociación entre la memoria de trabajo (MT) y la cantidad de notas y su relación con el rendimiento de la interpretación consecutiva (IC) para evaluar su eficiencia predictiva para las pruebas de aptitud en IC. Para seguir los objetivos del estudio, se administraron dos tareas de WM y una tarea de CI a 30 estudiantes de traducción de máster. Los resultados indicaron una relación positiva y significativa entre una de las medidas de WM (rango de lectura) y la cantidad de notas. Además, hubo una relación significativa entre ambas medidas de rendimiento de WM y CI, y también entre la cantidad de medidas y el rendimiento de CI. Además, los resultados del análisis de regresión indicaron que ambas medidas de WM eran predictores del desempeño de IC, mientras que la variable “cantidad de medida”; no era un predictor. Con base en los resultados, se demuestra además que WM es un componente eficiente en las pruebas de aptitud, mientras que la cantidad de medidas no cumplió completamente con los criterios y fue rechazada como un criterio fiable y no pudo incluirse como un subcomponente en las pruebas de aptitud.

Palabras clave: Interpretación consecutiva (CI), memoria de trabajo (WM), toma de notas, cantidad de notas, prueba de aptitud
1. Introduction

Consecutive Interpreting (CI) is one of the modes of communication which enables people to exchange information in various settings (for example in a police station, in court, at a press conference). Pöchhacker (2004) defines CI as the rendition of a whole source text segment by segment, during which the interpreter can take notes. He distinguishes between two types of CI: CI with note-taking and short CI without notes. A high quality performance is absolutely critical in CI in order to avoid misunderstandings and create efficient communication between the parties or between the speaker and his/her audience.

Interpreting institutes try to select candidates who have a higher chance than others of becoming successful interpreters. Thus, most of them apply various aptitude tests to evaluate candidates for the program. However, aptitude has been rarely addressed in interpreting studies (IS) (Dastyar, 2018). Furthermore, there is neither consensus on the components of aptitude tests nor enough empirical studies focusing on the issue. Various schools apply different aptitude tests with various components (Timarová & Ungoed-Thomas, 2008). Moreover, there is scarcity of studies like those of Timarová & Ungoed-Thomas (2008) or Donovan (2003), which reported comprehensive information in terms of the various aptitudes and their components applied in interpreting schools. Therefore, this domain calls for further investigation in order to examine the appropriateness of the various components to be included in or excluded from the aptitude tests.

The main contribution of the present study in this particular field of research is the focus on specific parameters believed to determine performance in CI, in other words, working memory (WM) and note quantity. WM has not received a proper attention in the context of aptitude tests, and few studies have investigated the predictive value of WM (e.g. Lin, Lv, & Liang, 2018). Moreover, researchers have commonly focused on simultaneous interpreting (SI) (Dong & Cai, 2015), while there is no clear picture of a phenomenon in CI. In addition, the association between WM and note-taking in CI has not been investigated to date, and there are no prior studies with focus on note-taking, and its predicting capabilities in interpreting aptitude tests. Therefore, the present investigation examines the association of WM with note quantity as well as the relationship between WM and note quantity with CI performance in English-Persian CI. The aim is to evaluate the predictive capabilities of the variables for CI performance, and to contribute to the literature with regard to aptitude tests which can in turn improve CI performance.

Based on the reported results, two questions will be addressed in the proposed research and it is hoped that the results will enrich us with valuable knowledge in this domain. The first research question examines the association between WM capacity and note quantity in CI: 1) Is there any significant relationship between WM capacity
and note quantity in CI? The second question focuses on the predictive power of WM and note quantity in CI performance to evaluate their effectiveness for aptitude tests:

2) Are WM and note quantity predictors of CI performance?

2. Aptitude tests in interpreting

An aptitude test is a common method employed to evaluate certain abilities and skills of interpreting candidates in interpreter training institutions and universities. Aptitude tests have been defined in various ways. In educational psychology, Dastyar (2018) states that an aptitude test is referred to as a standardized test which is applied to measure or predict some abilities of individuals for performing a certain task. Therefore, aptitude can be applied at the beginning of a program (for example an interpreting program) in order to predict the degrees of success in candidates.

As Su (2015) argues, there are limited resources and methods for selecting qualified interpreting candidates, and furthermore theoretical findings are inconsistent in this domain. The empirical studies on aptitude tests have examined the efficiency and predictability of some variables, including L2 proficiency (Mayor, 2015); soft skills and personality traits (Bontempo & Napier, 2011; Lopez Gomez, Bajo Molina, Padilla Benitez, & Santiago de Torres, 2007; Rosiers, Eyckmans, & Bauwens, 2011); verbal fluency, and oral paraphrasing (Russo, 2014; Russo & Pippa, 2004; Skaaden, 1999); (working) memory and the related variables (Darò, 1995; Lopez Gomez et al., 2007). As can be seen, a diverse range of variables have been examined, and it is difficult to draw a general trend line in this domain. Furthermore, there are a lot of variables which have not yet been investigated, including ‘note quantity’. Moreover, the majority of studies have mainly been devoted to SI, and signed language interpreting, while CI has rarely been addressed, with few exceptions (e.g. Mayor, 2015).

There are some studies in this domain which provide us with an overview of admission tests, and their components which are applied by various institutes. For example, Timarová & Ungood-Thomas (2008) surveyed 18 schools in terms of the type, and content of their admission tests. They summarized the results and listed the tests in order of the number of schools which used them (see Table 1).

As can be seen, the majority of schools apply short consecutive tests (with or without note-taking); checking the applicant’s general knowledge is very common in various schools, and written translation, sight translation, and oral presentation are among the less frequent skills and abilities that are tested. As for WM and note-taking (the focus of the present study), memory is checked indirectly and few schools apply summary and written skills as a component of aptitudes; note-taking is not applied at all.

In this section, firstly, the current aptitude test (which is applied at the University of Bologna) is presented as an example, then one of the recent suggested models for interpreting aptitude is reviewed.
Table 1. Ranking of admission tests by number of schools using them by Timarová & Unggoed-Thomas (2008, p. 36)

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short consecutive C-A / B-A / A-B 2-5 mins. Mostly without notes, in some cases candidates are allowed to take notes but told to concentrate on memory.</td>
<td>16</td>
</tr>
<tr>
<td>Interview (mostly in A and/or B, but some cases of C), may focus specifically on areas of general knowledge.</td>
<td>10</td>
</tr>
<tr>
<td>General knowledge questions (written or oral)</td>
<td>9</td>
</tr>
<tr>
<td>Written translation A-B / B –A / C-A</td>
<td>6</td>
</tr>
<tr>
<td>Sight translation C-A / B-A / A-B (at discretion of jury in a number of cases)</td>
<td>6</td>
</tr>
<tr>
<td>Written summary, in active language, of written or oral speech in active language e.g. maximum 100 word abstract of 600 word text</td>
<td>5</td>
</tr>
<tr>
<td>Oral presentation in A or B. 2-5 minutes long, after 1-15 minutes preparation</td>
<td>4</td>
</tr>
<tr>
<td>Written essay in active language(s) 300-500 word</td>
<td>2</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>2</td>
</tr>
<tr>
<td>Written summary, in active language, of written or oral speech in passive language</td>
<td>1</td>
</tr>
<tr>
<td>Paraphrasing, gap-filling, punctuation, synonyms</td>
<td>1</td>
</tr>
<tr>
<td>Essay in passive language(s) 300-500 words</td>
<td>1</td>
</tr>
<tr>
<td>Oral presentation in C on current affairs</td>
<td>1</td>
</tr>
<tr>
<td>Oral analysis of metaphorical or ambiguous headlines</td>
<td>1</td>
</tr>
<tr>
<td>Shorter oral summary of long oral presentation</td>
<td>1</td>
</tr>
<tr>
<td>Oral presentation in C language on the basis of a short list of words that must be included</td>
<td>1</td>
</tr>
</tbody>
</table>

The admission exam which was administered until the academic year 2018/2019 at the University of Bologna was composed of three parts: a) A cloze test in the B and C languages of candidates, b) A recall for the candidates’ B and C languages, and c) Oral paraphrasing from Italian into Italian. Since the academic year 2019/2020 (to optimize the procedure), the new aptitude test was introduced which includes 3 recalls (Language A>B, B>A, and C>A) and an oral interview on current affairs and contemporary history. Recall is regarded as an important component in both versions (M. Russo, personal communication, April 6, 2020).

One of the recent models of aptitude tests in interpreting has been suggested by Gambrell (2018) (see table 2). The components are similar to the common aptitudes and include, a cloze test, synonyms, Wisconsin Card Sorting Test (WCST), dual-task training memory exercise, and interviews. Each component, measures certain abilities
and together with other components, a candidate’s potentials and future performances are evaluated. For example, the cloze test is a measure of B language proficiency; the dual-task training memory exercise evaluates the candidate’s ability to receive and produce simultaneously, etc. (Gambrell, 2018).

Table 2. The Proposed aptitude test model (Gambrell, 2018, p. 87)

<table>
<thead>
<tr>
<th>Test</th>
<th>Language Fluency</th>
<th>General Knowledge</th>
<th>Memory</th>
<th>Expressive Ability</th>
<th>Linguistic Skills</th>
<th>Personality</th>
<th>Stress/Anxiety</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloze</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synonyms</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td></td>
<td></td>
<td>x (problem solving)</td>
<td>x (cognitive flexibility)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual-task memory</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviews</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although there are similarities in the contents of interpreting aptitudes, to a large extent they are different; the universities or interpreting schools commonly have their own screening methods, and there is no consensus over certain aptitudes. Regardless of the differences, scholars have frequently included memory in aptitude tests either directly or indirectly. Various forms of memory tests have been examined in empirical studies or included in test batteries, including various components of the Wechsler memory test comprising Digit Span (Lambert, 1991; Lopez Gomez et al., 2007); recall test (Gerver, Longley, Long, & Lambert, 1989); short-term memory (Alexieva, 1993); and dual-task memory (Gambrell, 2018).

Considering the importance of WM in interpreting, it has rarely been investigated in the context of aptitude tests with a few exception including Lopez Gomez et al. (2007) who implemented Digit Span test (as a subcomponent of Wechsler). However, there are quite a number of researches which have been devoted to the association between WM and interpreting, which are reviewed in section 3.
As was pointed out earlier, there is no consensus over the content of the aptitude tests and the predictive efficiency of the components in interpreting, especially CI. According to the literature, WM is a prerequisite for interpreting which has been proved as an efficient predictor in certain studies, but there is still no clear picture of the phenomenon in aptitude tests. Moreover, note-taking is also among the pivotal skills in CI. However, neither the degree of its application nor its predictive capabilities for aptitude tests has been addressed to date. Therefore, the present study focuses on these variables to shed some light on this issue. The ultimate objective is to report the efficiency of the variables for aptitude tests of CI. The findings might be fruitful in efficient screening in interpreting programs.

3. WM and interpreting

CI is a cognitively demanding activity as compared to other human activities. WM is one of the cognitive functions underlying interpreting, and is regarded as a determining factor (e.g. Chmiel, 2016; Dong, Liu, & Cai, 2018; Wen & Dong, 2019).

The concept of WM was proposed by Baddeley and Hitch (1974) as a new concept for short-term memory. The later version of the multi-component model of WM was introduced by Baddeley (2000) who included four main components for WM: the central executive, the phonological loop, the visuospatial sketchpad, and the episodic buffer. Each component has a special role in storing, processing, and retrieving information.

As Timarova et al. (2014) put it, in general, studies on WM in the context of interpreting have mainly focused on two topics: a) comparison between interpreters and non-interpreters on WM or comparison between the interpreters of different levels: professionals, students, etc., and b) association between WM and interpreting performance.

As for the first category, mixed results have been reported, but they are in favor of the superiority of interpreters over non-interpreters (Babcock & Vallesi, 2017; Chincotta & Underwood, 1998; Chmiel, 2016; Christoffels, De Groot, & Kroll, 2006; Hiltunen, Pääkkönen, Vik, & Krause, 2016; Hiltunen & Vik 2017; Köpke & Nespoulos, 2006; Lin et al., 2018; Liu, 2001; Liu, Schallert, & Carroll, 2004; Padilla, Bajo, Cañas, & Padilla, 1995; Wen & Dong, 2019).

In the second category, a positive and significant relationship between a higher WM and better interpreting performance has been reported (Christoffels, De Groot, & Waldorp, 2003; Hodáková, 2009; Injoque-Ricle, Barreyro, Formoso, & Jaichenco, 2015; Khatib, 2003; Timarova et al., 2014; Tzou, Eslami, Chen, & Vaid, 2012) with the exception of few studies like that of Wang (2016) who reported a lack of such significant association between the variables in signed language interpreting.

Yet, the third category may be added in which the researchers have examined the effect of memory training on interpreting performance or the effect of interpreter...
training on memory (e.g. Chmiel, 2016; Dong, Liu, Cai, 2018; Yenkimaleki & van Heuven, 2017).

The majority of studies, as Dong and Cai (2015) state, have focused on SI rather than CI. Therefore, WM has not received a proper attention in the context of CI, and the effect of WM in CI is unclear. As for English-Persian language pair, only one study has been reported (Khatib, 2003). Khatib (2003) administered Paced Auditory Serial Addition Test (PASAT) as a measure of WM. However, a general WM cannot be measured by employing it. PASAT is frequently used to evaluate attentional functioning, and information processing (Tombaugh, 2006) while WM is a combination of storing and processing. Furthermore, PASAT appears to involve the storage and processing tasks of minimal information chunks (the digit), while WM is a combination of storing and processing longer chunks (a sentence, several digits, etc.).

Therefore, more studies are required to apply the simple and complex measures of WM in order to examine the association between WM and CI performance focusing on the Persian language. As Dong and Cai (2015) put it, considering the close similarities between the CI process and the storage-plus-processing definition of WM, devoting future study to examine the role of WM in CI as well as its effect on other variables like note-taking is promising. The results can be helpful in increasing the understanding of this domain and modifying aptitude tests for selecting candidates which in turn enhances the quality of CI.

4. Note-taking and CI

One of the pivotal skills in CI is note-taking, and as Chen (2017) put it, note-taking provides an opportunity for investigating the process of CI. Note-taking is among the main features of CI (Pöchhacker, 2004), and efficient note-taking is required for successful performance. Moreover, note-taking helps an interpreter avoid memory overload (Mahmoodzadeh, 1992), and thus plays a supportive role for memory (Yamada, 2018).

The empirical studies in the area of note-taking in CI have focused on various topics, and variables, that is the association between note quality/quantity and CI performance (Cardoen, 2012; Dai & Xu, 2007; Dam, 2007; Dam, Engberg, & Schjoldager, 2005; Her, 2001; Liu, 2010; Wang, Dandan, & Ling, 2010); the language choice and form of note-taking (Abuin Gonzalez, 2012; Dai & Xu, 2007; Dam, 2004a; Dam, 2004b; Dam, 2007; Lung, 2003; Marani & Heidari Tabrizi, 2017; Szabó, 2006; Wang et al., 2010); process of note-taking (Andres, 2002; Chen, 2108); and comparison of professionals and non-professionals on note-taking (Abuin Gonzalez, 2012).

As for the association between note quantity and CI performance, Dam et al. (2005), after a pilot study, reported that longer notes result in a better target text. Furthermore, Dam (2007) examined the note-taking of five professional interpreters interpreting from Spanish into Danish. She concluded that more notes lead to a better CI performance. Unlike Dam et al. (2005), and Dam (2007), the following study
found contradicting results. Cardoen (2012) with the participation of three subjects and by focusing on their notes and CI from Spanish into Dutch, found that the chunks with fewer notes were more fluent. In addition, Chen (2016) reviews the following studies supporting a lack of significant association between the note quantity and Chinese-English CI (Dai & Xu, 2007; Liu, 2010; Wang et al., 2010).

Based on what has been reported, the results are inconsistent, but tip the balance slightly in favor of a lack of positive and significant relationship between note quantity and CI performance however, there is a limited number of studies. Furthermore, previous studies were mainly carried out with a small number of participants and were devoted to western European, and Chinese languages with a few exceptions (Szabó, 2006 with a focus on Hungarian-English). Therefore, it is hard to come to a conclusion and generalization. Thus, the present investigation is carried out with 30 participants and is devoted to the English-Persian language pair which is untouched in this domain.

4.1. WM and note-taking in CI

Note-taking is a demanding activity which includes storing and manipulating information (Lorek, Centifanti, Lyons, & Thorley, 2019). According to Bui and Myerson (2014) WM is among the cognitive abilities that may have an effect on note-taking. As Chen (2016) states, there is scarcity of research on the association between cognitive abilities and note-taking in CI. Furthermore, the association between WM and note-taking in CI is untouched. However, the association of WM and note-taking in other contexts, such as lecture note-taking has been studied, and as Bui and Myerson (2014) put forward, mixed results have been reported in this domain: some studies have reported that there is a significant association between WM and lecture note-taking (e.g. Divesta & Gray, 1973; Hadwin, Kirby, & Woodhouse, 1999; Kiewra & Benton, 1988; Kiewra, Benton, & Lewis, 1987; McIntyre, 1992). While some other studies have reported a lack of such significant association (Cohn, Cohn, & Bradley, 1995, Peverly et al., 2013).

As pointed out earlier, the association between WM and note-taking in CI and the predictive capabilities of note-taking has not been investigated to date. Therefore, various studies in this field can help scholars draw conclusions regarding the associations between variables which may be helpful for CI trainers and practitioners. As such, the present study examined the relationship between WM and note-taking and the predictive power of note-taking with a view to help scholars modify the aptitude tests in order to improve CI training, and performance.
5. Method

5.1. Participants

30 Persian speaking (Persian was their L1) MA translation students (14 males and 16 females) studying at the University of Isfahan, aged 22-30 years, participated in this study. They had passed a 16-session course on interpreting during which they had acquired some theoretical and practical familiarity with interpreting. The content of the course may differ slightly from one class to the next. However, the professors commonly focus on theoretical aspects of interpreting as well as devoting few sessions to CI practice.

These participants were selected from among 50 students, who had a) passed the Oxford Placement Test (OPT) and got at least a minimum score in the proficiency level (C2=55) and b) got a score of less than 2.5 in the self-report questionnaire. The combination of the OPT and questionnaire results allowed the researcher to have a homogenous group. Therefore, the finally selected 30 participants as proficient bilinguals were the students of similar proficiency in English language and similar theoretical and practical familiarity with interpreting.

5.2. Tasks

5.2.1. WM tasks

(Auditory and Forward) Digit Span Test: This test is a simple span test which measures the verbal WM. In the later version of the test, which is administered via a computer, the test is verbally carried out; on each trial participants are presented with a series of digits, each of which is presented at one time. At the end of each test, the participants attempt to recall the digits in the order of their presentations and type them via key press. The digits start with two ones in the first series and end with 9 in the last one. After each successfully completed trial, the number of digits presented increases by one digit in each next trial. After a failed trial (i.e., in case any digits are missing and/or when they have a wrong order), the number of the presented digits remains the same for the next trial. The task ends for a participant who makes errors at two sequential trials in a given Digit Span. A Digit Span includes, the maximum number of digits correctly recalled.

The Persian version of the test developed by Khodadadi and Amani (2014) was employed in the current study. Since, the forward Digit Span is applied as a measure of WM capacity (Stone, 2017), and has been frequently applied in studying WM in the context of interpreting (e.g. Christoffels et al, 2003; Injoque-Ricle et al., 2015), it was decided to apply it as one of the WM measures. Furthermore, the Persian version of the software was available which facilitated the process of the research.
Reading Span Test: This test, which was devised by Daneman and Carpenter (1980), is a complex span test capable of measuring a general WM. Through this measurement, individual differences in WM capacity can be examined. As Daneman and Carpenter (1980) pointed out, the Reading Span significantly correlates with both reading and listening comprehension. The Persian version of this test (Khodadadi, Asad Zadeh, Kalantar Ghoreishi, & Amani, 2014), which has been developed and validated on the basis of Persian language criteria, was applied with an automatic scoring procedure in this research. In this test, both the storing and processing abilities were scored and summed to obtain the final score.

In the Reading Span test, a series of short sentences are presented on the screen. The sentences start with 2 ones in the first series and end with 7 in the last series. They increase by one sentence for each next series. After each series of sentences, a table is presented on the screen. The participants are expected to select two types of answers via key press: a) whether the sentences they have seen on the screen are true or false, and b) whether they can recall the last word of each sentence in the exact order.

5.2.2. CI task

A recorded video lecture of 4.48 minute in English was used as CI task, the topic of which was ‘Why should we learn a new language?’ The lecturer was a native American English speaker. The text did not have any technical terms and thus, knowledge of everyday language could suffice for the material interpretation. The source text included 702 words with a delivery rate of 146.25 WPM. After each short paragraph, the researcher paused the video to allow the participant finish interpreting and then continued the video. (see Appendix C). The participants were supposed to interpret the text from English into Persian after each pause. All of them were provided on paper with a pen for note-taking.

5.3. Scoring the interpreting tasks

The revised version of Carrol’s Scale by Tiselius (2009) was used as a rubric for scoring the interpreting tasks (see Appendices A and B). This rubric is holistic and has two components: intelligibility and informativeness. Because of the non-compositional nature of this rubric, scoring is easy and consistency is promoted during the scoring procedure. Three raters (PhD candidates in translation studies) scored the interpreting tasks independently. The final score of each participant was the average of three raters’ scores. High inter-rater reliability ($r = .897, p <.001$) ascertained the reliability of the scoring procedure.
5.4. Procedure

The participants took part in the data collection phase, one by one in a quiet classroom. Each participant first took the Digit Span and then the Reading Span, and finally the CI test. The WM tests were automatically administered and scored on a laptop in the pre-established order. For the CI test, the recorded video speech was played on the laptop. Each participant was supposed to interpret the content after each pause. The researcher recorded the entire interpreting tasks with a voice recorder for later analysis.

All recorded interpreting tasks were transcribed verbatim. One of the raters listened to the interpreting tasks through a high quality headphone, and transcribed them meticulously. In order to come to a reliable result, he listened to each interpreting task several times. Each transcribed interpreting task was then scored by the raters according to the revised version of Carrol’s scale by Tiselius (2009). Following the Tiselius method, each interpreting unit was printed on a separate page, with the interpreted rendition at the top and the original at the bottom. The intelligibility scale was at the very top of each page and the informativeness scale at the very bottom. The number of words in each note (note quantity) was calculated by the raters for further analysis; each word or combination of letters between two spaces was regarded as a unit of calculation.

5.5. Data analysis

In order to analyze the data, the SPSS software version 21 was applied. Because of the parametric nature of the data, a Pearson Correlation Coefficient, and the Multiple Linear Regression were conducted for the analysis. First, a Pearson Correlation was conducted to assess the association between: WM variables (Digit Span and Reading Span) and note quantity, WM variables and CI performance, and note quantity and CI performance. Second, to examine the predictive effect of ‘WM measures’ and the variable ‘note quantity’ on CI performance, a multiple linear regression analysis was conducted. During the analysis, both variables of WM and the variable ‘note quantity’ were used as input variables into the regression model. This was done because their association with CI was significant based on a Pearson correlation analysis.

In order to examine the multicollinearity among the independent variables for the regression model, the Variance Inflation Factor (VIF) was calculated; the VIF values for three independent variables were: 1.49, 2.84, and 2.18. Considering that all values were between 1-10, it was clear that multicollinearity was not present.
6. Results

Descriptive statistics for all variables of the study including Digit Span, Reading Span, Note Quantity, and CI performance are shown in Table 3.

Table 3. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI performance</td>
<td>9.78</td>
<td>1.74</td>
<td>6</td>
<td>12</td>
<td>-1.22</td>
<td>.141</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.93</td>
<td>1.25</td>
<td>5</td>
<td>10</td>
<td>-.64</td>
<td>.31</td>
</tr>
<tr>
<td>Reading Span</td>
<td>76.72</td>
<td>10.12</td>
<td>50</td>
<td>94.40</td>
<td>-.67</td>
<td>.72</td>
</tr>
<tr>
<td>Note-quantity</td>
<td>50</td>
<td>13.75</td>
<td>31</td>
<td>79</td>
<td>.767</td>
<td>-.332</td>
</tr>
</tbody>
</table>

The Pearson correlation analysis indicated that, there was a positive and significant relationship between Reading Span and the variable ‘note quantity’. However, there was not a significant association between the other measure of WM—that is, Digit Span and the variable ‘note quantity’. Moreover, there were positive and significant relationships between CI performance and both WM measures, and also between CI performance and the variable ‘note quantity’ (see Table 4).

Table 4. Pearson Correlations between Consecutive Interpreting Performance, Digit Span, Reading Span, and Note Quantity

<table>
<thead>
<tr>
<th></th>
<th>CI Performance</th>
<th>Digit Span</th>
<th>Reading Span</th>
<th>Note-quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI performance</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Digit Span</td>
<td>.643(<strong>).505(</strong>)</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reading Span</td>
<td>.880(**)</td>
<td>.505(**)</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Note-quantity</td>
<td>.573(**)</td>
<td>.165</td>
<td>.701(**)</td>
<td>1</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level.

A multiple linear regression model was tested, including CI performance as dependent variable and two WM tasks (Digit Span and Reading Span), and the variable ‘note quantity’ as independent variables. The initial model was statistically significant, $F(3, 26) = 41.457, p < .05, R^2 = .807$. The independent variables predicted the
dependent variable: Digit Span test \((\beta = .270; t = 2.270, p = .011)\) and Reading Span test \((\beta = .734; t = 5.349, p = .000)\). The variable ‘note quantity’ was not significant according to the regression analysis (see Table 5).

Table 5. Multiple Linear Regression Model

<table>
<thead>
<tr>
<th>Model 1</th>
<th>(\beta)</th>
<th>(T)</th>
<th>(p)</th>
<th>(R^2) adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.520</td>
<td>.018</td>
<td>.807</td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>.270</td>
<td>2.720</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>Reading Span</td>
<td>.734</td>
<td>5.349</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Note Quantity</td>
<td>.013</td>
<td>.109</td>
<td>.914</td>
<td></td>
</tr>
</tbody>
</table>

Note. Dependent variable: Consecutive Interpreting Performance

7. Discussion

The first aim of the present study was to answer the following question: ‘Is there any significant relationship between WM capacity and note quantity in CI?’ After examining the association of the variables, it was found that a positive and significant relationship between one of WM measures (Reading Span) and note quantity existed. Taking that into consideration, Reading Span is a complex test and is a measure of general WM and it may be concluded that interpreters with higher WM can take longer notes in CI. Unlike Reading Span, there was no association between Digit Span and note quantity. Digit Span is a simple span test and measures the phonological loop or verbal WM. Therefore, it may be implied that the phonological loop alone does not play a significant role in note quantity in CI.

The positive association between WM and note-taking was consistent with a number of studies (e.g. Hadwin et al., 1999; McIntyre, 1992), and inconsistent with certain others (e.g. Cohn et al., 1995; Peverly et al, 2013). It should be noted that all mentioned studies examined lecture note-taking not CI note-taking, and there have been no prior studies on the association of WM and note-taking in the context of CI. Therefore, the results cannot be further discussed. Based on the results of the present study and most of those carried out on lecture note-taking, it may be deduced that subjects with higher WM scores can take longer notes compared to subjects with lower WM scores. However, the association should be further investigated in CI to reach more reliable conclusions with more power for generalization.

The second aim of the investigation was to answer the following question: ‘Are WM and note quantity predictors of CI performance?’ The multiple linear regression
analysis indicated that both variables of WM could predict CI performance; Reading Span was more powerful as compared to Digit Span; this finding was in line with the results of a Pearson Correlation analysis where the value was more powerful for Reading Span compared to that of Digit Span. Note quantity failed to be a predictor of CI performance. In other words, WM was so strong that it could neutralize the effect of note quantity.

The value of a Pearson Correlation between note quantity and CI performance was positive and significant. This result is consistent with Dam et al. (2005), and Dam (2007) who reported a significant association between the variables stating that more notes lead to a better target text. Furthermore, the present result is in contrast with Cardoen (2012) who reported that the chunks with fewer notes were more fluent. Moreover, the findings do not support studies like those of Dai and Xu, 2007; Liu, 2010; Wang et al., 2010 who observed a lack of positive, and significant association between note quantity and CI performance. To put it simply, the results of the present investigation support a positive and significant association between note quantity and CI performance, but the association is not strong (r= .573). However, the present results reject the predictive value of note quantity. That is, note quantity is not sufficiently strong and reliable to be applied as a predictive component in CI performance or an aptitude test.

WM, especially its complex measures is sufficiently reliable for predicting CI outcomes. Both the correlation and regression values were significant for WM variables, especially the values of Reading Span. Reading Span is regarded as a complex test and a measure of general WM. Therefore, it can be concluded that interpreters with higher WM perform better in CI, and WM should be included in aptitude tests for CI. These results are in line with the general trend line in this domain, in other words, a positive association between WM and interpreting performance (Christoffels et al., 2003; Hodáková, 2009; Injoque-Ricle et al., 2015; Khatib, 2003). Additionally, the findings are consistent with the predictive capability of WM measures for interpreting, reported by some researchers. Lin et al. (2018), for example reported that Reading Span can positively predict SI fluency and Lopez Gomez et al. (2007) who concluded that cognitive abilities including Digit Span have predicting efficiency in signed language interpreting.

Considering the participants of the study as candidates for the interpreting program- N=30, homogeneous proficient bilinguals with a similar degree of familiarity with interpreting, the outcome is sufficiently reliable to be generalized to similar circumstances. However, strong generalization calls for further investigations with more participants, and also replication of the study in other language pairs. Moreover, generalization is not among the main objectives of the present study and providing some implications for aptitude tests are regarded as pivotal.
8. Conclusions

The present study examined the relationship between WM and note quantity and their impact on CI performance to answer two questions: a) 'Is there any significant relationship between WM capacity and note quantity in CI?', and b) 'Are WM and note quantity predictors of CI performance?'

As for the first question, it was found that there existed a positive and significant association between one of WM measures (Reading Span) and note quantity. This result is in line with Hadwin et al. (1999) and McIntyre (1992) who reported the significant association between the variables in lecture note-taking. In terms of the second question, the results indicated that WM was a predictor of CI performance while note quantity failed to be a predictor. The predictive efficiency of WM, which proved in the present study is in accordance with certain studies, including Lin et al. (2018) who reported that Reading Span can positively predict SI fluency and Lopez Gomez et al. (2007) who concluded that Digit Span has a predicting efficiency in signed language interpreting.

The general conclusions of the study can be stated as: a) there is a positive and significant association between WM and note quantity in CI and b) WM can be used as a reliable factor for predicting CI performance, and it can be included as a component in aptitude tests for interpreting programs while note quantity is not a reliable component for CI aptitude tests and predicting CI performance. In general, the findings of the present study further support the determining impact of WM and its predictive capabilities as well as corroborating the theoretical and empirical studies with similar findings.

The present study was among the first empirical investigations to focus on WM and note quantity in CI as well as being among the first in the English-Persian language combination. Therefore, in addition to the findings, the study itself paved the way for future studies and investigations to be carried out in this domain with a focus on English-Persian or other language pairs. It is hoped that the results could enrich researchers with valuable knowledge along these lines. Moreover, the results may lead to a better understanding of the phenomenon and increase the quality of CI by screening appropriate candidates for interpreting programs based on sound aptitude tests. Regardless of the limitations and findings, the present study opens up a new window for interpreting research that is, investigating the association between WM and CI note-taking which has been ignored to date.

The number of participants, which might have affected the results is one of the limitations of this study. This investigation was carried out with the participation of 30 students studying at the University of Isfahan. This number is statistically reliable, and is not unusual in interpreting researches. However, more participants, may lead to a better outcome with a higher mean power to generalize the findings. At the same time, it would have been more difficult to handle the investigation with more subjects.
The procedure of implementing, recording, transcribing, and scoring the tasks by three raters was a delicate and time consuming activity.

Furthermore, applying other measures of WM such as complex span tests including Operation Span would also contribute to knowledge acquisition on complex tasks such as CI. Studies with various durations between pauses may provide interesting results as well. Moreover, examining the predictive capabilities of note-taking and its components (form, quantity, quality, language) are suggested as promising domain for further investigation.

9. References


• Di Vesta, Francis J. and Gray, G. Susan (1973). Listening and note taking: II. Immediate and delayed recall as functions of variations in thematic continuity, note taking, and length of listening-review intervals. *Journal of Educational Psychology* 64 (3), 278.


- Donovan, Clare (2003). Entrance exam testing for Conference Interpretation Courses: How important is it?. *FORUM. Revue internationale d’interprétation et de traduction/International Journal of Interpretation and Translation* 1 (2), 17-44.


- Donovan, Clare (2003). Entrance exam testing for Conference Interpretation Courses: How important is it?. *FORUM. Revue internationale d’interprétation et de traduction/International Journal of Interpretation and Translation* 1 (2), 17-44.


• Lin, Yumeng, Lv, Qianxi and Liang, Junying (2018). Predicting fluency with language proficiency, working memory, and directionality in simultaneous interpreting. *Frontiers in psychology* 9, 1543.


Appendix A

Adapted Carroll’s Scales (Tiselius 2009)

<table>
<thead>
<tr>
<th>Scale of Intelligibility</th>
<th>Scale of Informativeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. The rendition is perfectly clear and intelligible. Like ordinary spoken Persian with few if any stylistic infelicities.</td>
<td>6. Reading the original changes the whole understood meaning. (6 should be given when reading the original completely changes the meaning that the rendition gave).</td>
</tr>
<tr>
<td>5. Generally clear and intelligible but with minor grammatical or stylistic peculiarities or unusual word choices, nothing that hampers the understanding.</td>
<td>5. Reading the original clarifies the understood meaning. The original’s differences in syntax, words and phrases alter the listener’s impression of the meaning to some extent.</td>
</tr>
<tr>
<td>4. The general idea is intelligible, but full comprehension is interfered with by poor word choice, poor style, unusual words and incorrect grammar. The Addressee will have to make an effort to understand the utterance.</td>
<td>4. Reading the original gives some additional information about syntax and words. It can also clarify minor misunderstandings in the rendition.</td>
</tr>
<tr>
<td>3. Masquerades as an intelligible utterance, but is actually more unintelligible than intelligible. Nevertheless, the idea can still be comprehended. Word choices, syntactic arrangements, and expressions are generally unusual and words crucial to understanding have been left out.</td>
<td>3. By correcting one or two meanings, mainly on word level, the reading of the original gives only a minor difference in meaning compared to the rendition.</td>
</tr>
<tr>
<td>2. Almost completely unintelligible. Although it does not seem completely nonsensical and the Addressee may, with great effort, discern some meaning.</td>
<td>2. No new meaning is added through reading the original, neither at the word level nor at the grammatical level, but the Addressee is somewhat more confident that s/he really comprehends the meaning intended.</td>
</tr>
<tr>
<td>1. Totally unintelligible and completely without meaning</td>
<td>1. No new meaning is added by the original, nor is the Addressee’s understanding of the rendition increased.</td>
</tr>
<tr>
<td>0. The original contains less information than the rendition.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Scales of intelligibility and informativeness on grading sheet

1. Scale of intelligibility on grading sheet

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Totally unintelligible</td>
<td></td>
</tr>
<tr>
<td>2. Generally unintelligible</td>
<td></td>
</tr>
<tr>
<td>3. Seems intelligible</td>
<td></td>
</tr>
<tr>
<td>4. General idea intelligible</td>
<td></td>
</tr>
<tr>
<td>5. Generally intelligible</td>
<td></td>
</tr>
<tr>
<td>6. Completely intelligible</td>
<td></td>
</tr>
</tbody>
</table>

2. Scale of informativeness on grading sheet

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Original contains less information than rendition.</td>
<td></td>
</tr>
<tr>
<td>1. Without any new information.</td>
<td></td>
</tr>
<tr>
<td>2. Not new information, strengthens the intended meaning.</td>
<td></td>
</tr>
<tr>
<td>3. Minor changes in meaning.</td>
<td></td>
</tr>
<tr>
<td>5. Original explains and improves.</td>
<td></td>
</tr>
<tr>
<td>6. Only new information.</td>
<td></td>
</tr>
</tbody>
</table>

Appendix C

The Transcribed Version of the Interpreting Task/ Video Lecture with the location of Pauses for Interpreting

The language I’m speaking right now is on its way to becoming the world’s universal language, for better or for worse. Let’s face it, it’s the language of the internet, it’s the language of finance, it’s the language of air traffic control, of popular music, diplomacy -- English is everywhere. (pause 1 for interpreting)

Now, Mandarin Chinese is spoken by more people, but more Chinese people are learning English than English speakers are learning Chinese. Last I heard, there are two dozen universities in China right now teaching all in English. English is taking over. (pause 2 for interpreting)

And in addition to that, it’s been predicted that at the end of the century almost all of the languages that exist now -- there are about 6,000 -- will no longer be spoken. There will only be some hundreds left. And on top of that, it’s at the point where instant translation of live speech is not only possible, but it gets better every year. (pause 3 for interpreting)

The reason I’m reciting those things to you is because I can tell that we’re getting to the point where a question is going to start being asked, which is: Why should we learn foreign languages -- other than if English happens to be foreign to one? Why bother to learn another one when it’s getting to the point where almost everybody in the world will be able to communicate in one? (pause 4 for interpreting)

I think there are a lot of reasons, but I first want to address the one that you’re probably most likely to have heard of, because actually it’s more dangerous than you might think. And that is the idea that a language channels your thoughts, that the vocabulary and the grammar of different languages gives everybody a different kind of acid trip,
so to speak. That is a marvelously enticing idea, but it’s kind of fraught. (pause 5 for interpreting)

So it’s not that it’s untrue completely. So for example, in French and Spanish the word for table is, for some reason, marked as feminine. So, “la table,” “la mesa,” you just have to deal with it. It has been shown that if you are a speaker of one of those languages and you happen to be asked how you would imagine a table talking, then much more often than could possibly be an accident, a French or a Spanish speaker says that the table would talk with a high and feminine voice. So if you’re French or Spanish, to you, a table is kind of a girl, as opposed to if you are an English speaker. (pause 6 for interpreting)

It’s hard not to love data like that, and many people will tell you that that means that there’s a worldview that you have if you speak one of those languages. But you have to watch out, because imagine if somebody put us under the microscope, the us being those of us who speak English natively. What is the worldview from English? (pause 7 for interpreting)

So for example, let’s take an English speaker. Up on the screen, that is, Bono. He speaks English. I presume he has a worldview. Now, that is Donald Trump. In his way, he speaks English as well. (pause 8 for interpreting)

And here is Ms. Kardashian, and she is an English speaker, too. So here are three speakers of the English language. What worldview do those three people have in common? What worldview is shaped through the English language that unites them? It’s a highly fraught concept. And so gradual consensus is becoming that language can shape thought, but it tends to be in rather darling, obscure psychological flutters. It’s not a matter of giving you a different pair of glasses on the world. (pause 9 for interpreting)

Now, if that’s the case, then why learn languages? If it isn’t going to change the way you think, what would the other reasons be? There are some. One of them is that if you want to imbibe a culture, if you want to drink it in, if you want to become part of it, then whether or not the language channels the culture -- and that seems doubtful -- if you want to imbibe the culture, you have to control to some degree the language that the culture happens to be conducted in. There’s no other way. (pause 10 for interpreting)