A Longitudinal Study of the Evolution of Working Memory amongst Interpreting and Translation Students at the FTI, University of Geneva

STEAD, Andrew, TRIPIER, Coralie

Abstract

This study developed and tested the idea that training in simultaneous interpreting leads to the working memory advantage generally observed among interpreters. Answering the question of the working memory advantage possessed by interpreters, and its causal link with interpreting training, is of fundamental importance in the study of simultaneous interpreting. Very few studies in the field have been able to carry out a longitudinal study to test the evolution of working memory in interpreters. We conducted a longitudinal study consisting of two sessions of experiments designed to test working memory, in interpreting students at the University of Geneva and in a control group of translation students. Contrary to our expectations we did not observe significant changes in either test group, nor the expected working memory advantage. The results of our experiments, possible reasons for the unexpected outcome and areas to focus on in the future are discussed.

Reference


Available at:
http://archive-ouverte.unige.ch/unige:88101

Disclaimer: layout of this document may differ from the published version.
ANDREW STEAD
CORALIE TRIPIER

A Longitudinal Study of the Evolution of Working Memory amongst Interpreting and Translation Students at the FTI, University of Geneva

Mémoire présenté à la Faculté de Traduction et d’Interprétation
Pour l’obtention du MA en Interprétation de Conférence
Directeur de mémoire : Prof. Kilian Seeber
Juré : Dr Alexis Hervais-Adelman

May 2016
STUDENT INFORMATION:

Andrew Stead

Coralie Tripier

Faculté de Traduction et d'Interprétation

University of Geneva

40, boulevard du Pont-d'Arve,

CH-1211 Genève 4, Switzerland

andrew.stead@etu.unige.ch

coralie.tripier@etu.unige.ch
Acknowledgements

We would like to extend our heartfelt thanks to Dr Alexis Hervais-Adelman for the tremendous support and guidance he has given us over these past 18 months. It was his inspiring classes that gave us the idea to investigate this area of research in the beginning.

We would like to also thank Dr Kilian Seeber for allowing us to carry out the research, and for the additional assistance he has given us throughout our journey.

We would like to thank all of our fellow interpreting students who contributed to this research through taking part in our experiments as test subjects, as well as the students of translation who joined us. Without all of you, this research truly would not have been possible.

We would like to extend special thanks to Loan Walter, our long suffering house-mate, who has both endured our dinnertime conversations about working memory and assisted in the recruitment of participants.

Finally, we would like to thank John Stead, Andrew’s father, for having proofread this paper as well having acted as a sounding post for ideas and problems throughout the research.
CONTENTS

ABSTRACT ..........................................................................................................................................................

1. OVERVIEW ...................................................................................................................................................

    1.1 NOMENCLATURE
    1.2 EXECUTIVE FUNCTIONS
    1.3 WORKING MEMORY
    1.4 BILINGUALISM AND WORKING MEMORY
    1.5 SIMULTANEOUS INTERPRETING AND WORKING MEMORY

2. WORK TO DATE ...............................................................................................................................................

    2.1 NATURE OF THE RELATIONSHIP

3. RESEARCH ......................................................................................................................................................

    3.1 RESEARCH QUESTION
        3.1 FACULTY OF TRANSLATION AND INTERPRETING AT THE UNIVERSITY OF GENEVA
    3.2 METHOD
        3.2A PARTICIPANTS
        3.2B PROCEDURE
            3.2B1 APPARATUS
            3.2BII D-SPAN
            3.2BIII O-SPAN
            3.2BIV DUAL N-BACK

4. RESULTS ...........................................................................................................................................................

    4.1 D-SPAN
    4.2 O-SPAN
    4.3 N-BACK

5. DISCUSSION AND CONCLUSION ......................................................................................................................

6. LIMITATION AND FUTURE RECOMMENDATIONS ..........................................................................................

8. BIBLIOGRAPHY ..................................................................................................................................................

APPENDICES ............................................................................................................................................................
ABSTRACT

This study developed and tested the idea that training in simultaneous interpreting leads to the working memory advantage generally observed among interpreters. Answering the question of the working memory advantage possessed by interpreters, and its causal link with interpreting training, is of fundamental importance in the study of simultaneous interpreting. Very few studies in the field have been able to carry out a longitudinal study to test the evolution of working memory in interpreters. We conducted a longitudinal study consisting of two sessions of experiments designed to test working memory, in interpreting students at the University of Geneva and in a control group of translation students. Contrary to our expectations we did not observe significant changes in either test group, nor the expected working memory advantage. The results of our experiments, possible reasons for the unexpected outcome and areas to focus on in the future are discussed. We conclude that it may be necessary to conduct similar tests on larger populations in the future in order to obtain more reliable data. This thesis contributes to an area of research with relatively few published papers and provides support to the design of future tests in the field.
1. INTRODUCTION

Simultaneous Interpreting (SI) is the process of conveying a message from a speaker to an audience, who do not share the same language, with little delay. Simultaneous interpreters work in conference settings, generally in soundproofed booths. SI requires simultaneous listening and processing of the source language, and formulation into the target language, as well as monitoring of output. It is therefore generally considered a challenging cognitive task. This view is propounded by Barik (1973); Gerver (1974); Gile (1995); Moser-Mercer, (1997) and Hervais-Adelman, Moser-Mercer, & Golestani, (2015) among others. Speakers deliver their speech into a microphone, and using the headset the interpreter renders the message into another language essentially simultaneously. Audience members may listen to the interpretation using headsets in the conference rooms, and may able to choose from a number of languages provided by different interpreters in separate booths (European Commission, 2014).

1.1 Nomenclature

The nomenclature in the field of SI research continues to develop, at the moment several terms are used synonymously to designate some processes and functions that are related to SI and to other fields of research. In this paper we will refer to:

- Executive Functions
- Working Memory
- Simultaneous Interpreting

1.2 Executive Functions

Executive functions (sometimes referred to as cognitive control), as described by Miyake et al., (2000) are:

“General-purpose control mechanisms that modulate the operation of various cognitive sub-processes and thereby regulate the dynamics of human cognition” (p. 50)

Four such executive functions are commonly proposed within the field, these are: Working Memory, Response Selection, Response Inhibition and Task-Set switching (Lenartowicz,
Kalar, Congdon, & Poldrack, 2010) although the exact terminology employed to describe these executive functions varies in the literature. These four domains of executive function are thought to be performed by the frontal lobe area in the brain. There is still some contention however over whether the executive functions are in some way linked and complementary to each other, or in fact separate and independent. One issue that arises in the study of executive functions is that the functions themselves cannot be studied directly, but only through the performance of tasks thought to tap those functions. Miyake et al explain an added dilemma of this approach, the “task-impurity problem” (2000, p. 90). Whilst experiments do require executive functions in order to be performed, they also require secondary cognitive processes which can be difficult to control for. They cite this as a possible reason for the continuing uncertainty over the link (or lack thereof) between the executive functions.

1.3 Working Memory

Working memory (WM) is commonly accepted within the available literature as an executive function. Baddeley and Hitch’s 1974 model was proposed to replace the previous concept of Short-Term-Memory (STM), and Baddeley (2000) explains that the Atkinson and Shiffrin model (1968) had three weaknesses.

Namely, difficulties “in accounting for the relationship between type of encoding and LTM (Long-Term-Memory) (Craik & Lockhart, 1972), (2) in explaining why patients with grossly defective STM had apparently normal LTM, and (3) in accounting for the effects of a range of concurrent tasks on learning, comprehending and reasoning.” (Baddeley, 2000, p. 418)

Baddeley and Hitch’s 1974 model of WM – later revised into Baddeley’s 2000 model - has since become the framework for several theories about working memory. The original model posits a “central executive” as a control unit for two “slave” systems, the phonological loop, and the visuospatial sketchpad (see Figure 1).

The phonological loop provides both storage for auditory data (the phonological store) through the unspoken rehearsal of words and an “inner ear” (the subvocal rehearsal component) allowing for the transformation of written data or images into phonological information (Baddeley, 2000). This component is assumed to be phonological in nature as opposed to semantic or visual.
More recently changes to the Baddeley and Hitch's model have been proposed to include an “episodic buffer” to address certain problems with the original model. These were issues with the memory of touch and smell. The episodic buffer is assumed to be an intermediary between both the central executive and its two slave systems, and LTM, allowing for the store and recollection of memory “episodes”. Baddeley (2000) further posits that whilst the episodic buffer is not likely to be based exclusively in one single area of the brain, it is also likely to be primarily housed in the frontal lobes.

![Diagram of Baddeley and Hitch's revised model](image)

1.4 Bilingualism and Working Memory

The effect of personal experience on cognitive performance has been a field of interest for many years, and there is now a general consensus that the brain can be shaped by it. Functional neuroplasticity (the ability of the brain to evolve and adapt to better perform some activities) has for example been found in people who regularly juggle, play video games, drive taxis or undergo musical training (Bialystok, Craik, & Luk, 2012). It is particularly interesting to note that London taxi drivers, who have an extensive knowledge of a complex road network, show greater grey matter volume in the posterior hippocampi (Maguire et al., 2000). In addition, the more they navigate through the city, the bigger their posterior hippocampi grow, suggesting a causal link between the task and the growth.

For a time, early multilingualism was thought to be detrimental to children's cognitive abilities. Goodenough (1926) compared studies of the children of immigrant parents in the
United States who did not speak English at home, he suggested that, “the use of a foreign language in the home is one of the chief factors in producing mental retardation as measured by intelligence tests” (p. 393). Later Grosjean (1982) took on this misapprehension, illustrating the extent of the misconception by citing Adler (1977) who described bilingual children as having split minds, being “neither here nor there”, marginal people. When describing a bilingual child, Adler writes “His standards are split, he becomes more inarticulate than one would expect of one who can express himself in two languages, his emotions are more instinctive, in short, bilingualism can lead to a split personality and, at worst, to schizophrenia" (p. 40). The first major study to contradict this assumption tested a group of monolingual children and a group of bilingual children on various memory experiments and showed a cognitive advantage for bilingual children on most tests, in particular those requiring symbol manipulation and reorganization (Peal & Lambert, 1962). Since then, being a multilingual individual has been regularly linked to enhanced cognitive performances (Bialystok et al., 2012; Bialystok, 2009, 2011; Diamond, 2010; Hilchey & Klein, 2011; Krashen, 2010). The executive functions in particular are thought to be affected by multilingual language production (Bialystok, 2009). Indeed, multilingual language production requires the constant involvement of the executive functions in order to direct attention towards the target language. This ongoing exercise of cognitive control in the linguistic domain is thought to be responsible for the difference in the executive functions of multilinguals compared to monolinguals.

As WM is an executive function, a global enhancement in the executive functions could entail an enhancement of WM (Bialystok, 2009). Furthermore, the majority of published experiments on WM and executive control of multilinguals show that these individuals generally tend to have poorer results in verbal recall and verbal fluency, showing a vocabulary deficiency compared to monolinguals, but better results in executive control (Bialystok et al., 2012; Bialystok, 2009).

If multilinguals show better results in executive control, it would be interesting to see whether training in SI, which could be considered to be a form of “extreme multilingualism” can further improve these results. Indeed, simultaneous interpreters exercise language control differently to other multilinguals, the latter having to use one language whilst suppressing another (or several) and the former having to divide attention between two languages, in order to monitor both the source speech and output. (Hervais-Adelman et al., 2015)
1.5 Simultaneous Interpreting and Working Memory

The link between WM and SI has been investigated by many researchers. It is believed that WM plays a crucial role in the performance of SI (Darò & Fabbro, 1994; Gerver, 1976; Moser, 1978), although the exact nature of this link remains open for discussion. Some experimental studies fail to find significant WM advantage in SI populations (Liu, Schallert, & Carroll, 2004; Nordet & Voegtlin, 1998). However, the majority of published studies do find a WM advantage amongst simultaneous interpreters (Babcock, 2015; Chincotta & Underwood, 1998; Christoffels, De Groot, & Kroll, 2006; Köpke & Nespoulous, 2006; Köpke & Signorelli, 2012; Signorelli, 2008; Timarová et al., 2014; Tzou, Eslami, Chen, & Vaid, 2011). We are inclined to work from the basis that interpreters show a general WM advantage compared to other groups.

2. WORK TO DATE

Work carried out until now has compared interpreters of differing degrees of experience with control groups composed of either students or bilinguals. Most of these studies have shown that bilingualism alone is not the causal factor for WM advantage in simultaneous interpreters.

Christoffels, De Groot & Kroll (2006) compared both the linguistic abilities and WM of Dutch native interpreters, bilingual university students (not studying interpreting) and native Dutch teachers of English as a foreign language. The interpreters outperformed students in both WM and linguistic ability. Interestingly, the interpreting group outperformed the teacher group in WM tests but not in linguistic ability tests. They conclude that WM is a “critical subskill” for SI.

Köpke and Nespoulous (2006) looked at the WM capacity of two groups of interpreters, novice and professional, with large mean age differences. They compared these interpreters to two control groups made up of student and multilingual subjects and found that tests focusing on the functions of the central executive showed an interpreter advantage, and that novice interpreters showed higher WM capacity than experts. The researchers suggest that expertise lessens the need to rely on WM, but that WM is nevertheless critical for interpreting at beginner level.

Timarová et al. (2014) attempt to concretely establish the link between simultaneous interpreting and executive functions. 28 practising interpreters working in EU institutions in Brussels were tested for interference resistance and response inhibition, their attention
shifting ability was measured, as was participants’ aptitude in “updating” (evaluation of incoming information followed by necessary changes to memory content with a view to completing the task). With reference to updating, Timarová et al. (2014) state:

“The resemblance to the demands of simultaneous interpreting can hardly be overstated: a continuous stream of incoming information needs to be retained briefly while it is being processed, and then “flushed” to make room for new information.” (2014, p. 145)

“Updating” was tested using a 2-back task. The participants were then given interpreting exercises to complete and a series of elements were measured, namely: lexical, semantic and syntactic processing. Global processing measures such as ear-voice span, active vocabulary and the speed of source text were also taken into account. Timarová et al. (2014) report that improved performance in the WM tasks correlates positively with some performance indicators in interpreting. Performance in the 2-back task showed a positive correlation with the correct interpretation of numbers. These results suggest that interference and response inhibition improve with increased experience in interpreting. The other two central executive functions, attention shifting and updating are found to be closely linked with performance in interpreting but do not seem to be modified by experience. The researchers consider the limits of their approach, as only quantitative measurements are made (as compared to qualitative assessments of interpreting performance), and the limited study size means that weaker correlations are less likely to be obvious. The researchers' approach is descriptive rather than explanatory. The research was also carried out extremely methodically, as can be seen in the paper: even the smallest considerations were taken into account to improve scientific rigour. Each decision about the research process is documented, and each parameter (even down to the participants' distance from the computer screen) is explained. Timarová et al. point to the need for further study into the link between experience in interpreting and executive functions.

Tzou et al. (2011) compared the WM and reading span1 of 11 first-year and 9 second-year students of interpreting with that of 16 untrained bilingual controls. All of the subjects were Mandarin native speakers. While the researchers found that the second-year students slightly outperformed the first-year students in WM performance, they failed to find a significant enough difference to draw any conclusion. However, they did find significant WM and reading span advantages in students of interpreting compared to untrained bilinguals.

---

1 Reading Span is a test developed by Grant, Daneman, & Carpenter, (1980). It is a measure of the ability to recall words at the end of a series of sentences. Subjects read sentences aloud and are asked to recall the final word of each sentence. The number of sentences presented to subjects increases incrementally, until a failure of three successive words is found. The maximum number of words recalled is deemed to be the reading span.
This result may point towards a WM effect due to SI training, but would need a longitudinal study to be confirmed. The researchers indeed indicate that a preliminary WM span assessment of the three groups would have allowed them to understand the direction of the link between WM span and SI training.

Another result arising from their study is the fact that WM span also turned out to be higher in individuals that showed greater second language proficiency. The researchers leave the following question open for discussion: could language proficiency explain the differences observed in SI performance and WM? Could training in SI enhance language proficiency skills rather than WM directly?

While their initial results regarding higher WM span in students of interpreting is to be taken into account in any study focusing on the impact of SI training on WM, it is of our view that the methodology they used was not always consistent. Indeed, the researchers had both students in interpreting and untrained bilingual controls undertake a battery of tests, which included SI exercises. Since the untrained bilingual controls have never followed any kind of training in SI, we do not believe that it is fair to assess their potential SI performance. This will influence the design of our tests, to be explained later.

Chincotta and Underwood (1998) carried out a research on bilingual digit span and articulatory suppression among experienced simultaneous interpreters and undergraduate student controls. 12 simultaneous interpreters and 12 student controls were tested for digit span with verbal recall, in both English (foreign language) and Finnish (native language). The digits were presented to the participants in Arabic numeral format and recall was marked by a bilingual examiner listening to the oral recall in both English and Finnish. Digit span was also tested under articulatory suppression conditions. In this mode, the participants uttered "la-la" for a period of 4 seconds before and 4 seconds after reading the numbers aloud, before proceeding to oral recall.

Both groups showed greater digit spans when recalling numbers in their mother tongue, but no significant differences were found between the groups in the control phase (basic digit span). However, the tests found an SI advantage under articulatory suppression conditions. The research indicates that the simultaneous interpreters group had learnt to counteract the effects of articulatory suppression through practice. The researchers, however, highlight that the lack of significant difference found under control conditions could be due to the small sample size, and that a larger study might yield a significant result. The researchers claim this by pointing towards the “insignificant” tendency of the SI group to show advantage under control conditions. A larger study would allow for this theory to be tested.
Nordet and Voegtlin (1998) asked the following question: do interpreters demonstrate better WM than the average population? If so, should the entrance exam to the interpreting program at the Faculté de Traduction et d’Interprétation (FTI) of the University of Geneva (Formerly École de Traduction et d’Interprétation, ETI) include a memory test in order to better identify the subjects that are more likely to become successful interpreters?

To answer that question, the researchers selected three groups: a group of 6 professional interpreters, a group of 7 interpreting students, and a group made up of 22 psychology students. All the subjects were native French speakers and resided in Geneva.

Nordet and Voegtlin chose to assess their subjects through a single test: the listening span. They hypothesised that WM evolved along with the practice of SI, and that hence, the professional interpreters should outperform the two student groups.

In the listening span, the subjects must firstly listen to a sentence and then evaluate if it is coherent or not (e.g. "monkeys lay eggs": not coherent, “lung cancer can be linked to smoking": coherent), and then memorize the last word. At the end of a series of sentences (varying in numbers), they must repeat all the last words in the order they appeared.

While the professional interpreters slightly outperformed the two other groups on all variables, the difference was not statistically significant. Their initial hypothesis was therefore uncorroborated.

According to the researchers, the lack of difference can be explained by the fact that simultaneous interpreters do not have a better WM than average, but rather use chunking strategies that allow them to expand their WM capacities. Moser (1978) considers that, expert simultaneous interpreters make more use of their anticipation capacities, thus allowing them to use fewer of their WM resources. According to Nordet and Voegtlin, since the words to be memorized were not always part of a coherent context, the interpreters could not use their chunking strategy to link them together.

It would have been interesting to know whether the coherent sentences’ last words were better retained by interpreters than the incoherent ones. However, unfortunately, due to difficulties encountered during the statistical process, the researchers were not able to report this result which could have supported their suggestions.

The researchers conclude that interpreters may not possess superior ability in WM but instead rely on a series of strategies that ultimately give an advantage in memory under normal conditions, but not under test conditions. They therefore recommend that the ETI's
entrance tests should not include a memory test, but rather an evaluation of a “strategies assimilation capacity” that is, an ability to adopt such strategies that are required as an interpreter (e.g. chunking, phonological rehearsal, generation of mental images), which we would argue is unfeasible over the period of a short entrance test.

Moreover, we are of the opinion that the study did not objectively assess the groups' WM. First of all, the groups were of very uneven size, and differed substantially in age (on average 35.8 years old for professional interpreters, 26.3 for interpreting students, and 22.6 for psychology students). The professional interpreters and interpreting student populations were also quite small. Secondly, the assessment of the sentence coherence relied on general knowledge, levels of which can vary according to the age and professional experience. For example, sentences such as "Nazi gold constituted a great controversy" (answer: coherent) or "French children wear uniforms to school" (answer: not coherent) do not set all subjects on an equal footing. This can in turn hinder the subjects' memory capacities due to a processing overload.

Finally, we do not think that a single test can provide a solid enough basis to draw any conclusion. In this case, the test used is a listening span, which is based on linguistic stimuli, only it would have been interesting to compare the subject's listening span to their digit span or other measures.

In conclusion, we think that the study, which contradicts the majority of the literature in this field, has not assessed extensively enough its subjects' WM. We would argue that in order to properly assess the WM capacity of individuals, it is necessary to perform more than one test.

Liu et al., (2004) examined whether differences in performance exist between professional simultaneous interpreters and individuals studying interpreting. 11 professional interpreters, 11 advanced interpreting students (end of second year) and 11 beginner interpreting students (end of first year), were tested. All participants were native speakers of Mandarin and all tests were conducted in English. The English level of the participants was established by their entry into the interpreting programmes from which they were recruited; in Taiwan and California, indicating that the subjects had obtained a TOEFL score of 600 or above, or a similar grade in a comparable test. Participants firstly underwent a WM test in the form of a listening span similar to that carried out by Nordet and Voegtlin. The listening span test required subjects to listen to a series of sentences, some of which were intentionally nonsensical, and decide whether each sentence was coherent or not. At the end of a series of sentences, the participants were asked to recall the final word of each sentence.
The second part of the test required the participants to interpret three speeches, all based on the subject of translation and interpreting, in order to ensure the presence a comparable level general subject knowledge in the three studied groups. The three speeches varied in difficulty and subjects were graded on their ability to interpret them.

The results showed that in the listening span test, no group; neither beginner, nor advanced students, nor professional interpreters held an advantage. However, and unsurprisingly, the professional interpreters did show improved performance in interpreting compared to the two student populations.

This is one of the few published studies that posits that professional interpreters do not hold a WM advantage over other groups. The study claims that differences in general WM capacity cannot be reliably inferred from the varying degrees of interpreting expertise. This is very interesting, as extremely few studies show this to be the case. One explanation for this uncommon result may be the small study size, as suggested by the authors. However, the study size in this test was not unlike other tested simultaneous interpreter populations [García (2014) in a review of relevant literature cites only one study with a number of participants greater than 30]. Another view, and the reason we are inclined to believe, is that the listening span test alone may not have been sufficient to establish whether or not a difference in WM capacity between the groups existed. It is our view that in order to demonstrate the presence (or lack) of a difference in WM capacity, a more extensive battery of tests should have been employed, as is usually the case of studies on WM and simultaneous interpreting. It is interesting to note that in the two studies that find that interpreters do not hold a WM advantage, both used the listening span as a measure of WM, and solely this test.

2.1 Nature of the Relationship

We consider that there is sufficient evidence available to sustain the view that interpreters have an advantage in WM, compared to other language professionals. What remains unclear is the nature of the relationship between WM development and SI. This has led to the question;

Does the WM advantage found in simultaneous interpreters arise during training, or is it present beforehand?

The need for a longitudinal study looking at the developmental relationship between WM and SI training has, in recent years, been pointed to by a number of scholars (Christoffels et al.,
A longitudinal study was finally carried out by Babcock (2015) to answer the following question: are interpreters "expert bilinguals", i.e. with the same cognitive abilities as other multilinguals, only at a higher level, or are they "unique bilinguals", with their own unique set of abilities?

To answer this question, she studied three groups: a group composed of 22 interpreting students in the European Master in Conference Interpreting programme (EMCI), a control group composed of 21 translation students and a second control group of 19 monolingual students of other subjects, in order to separate any potential effect of further language acquisition on the results. All students were enrolled at the University of Trieste, Italy. At the end of the second testing session, which took place at the end of the interpreting programme, two years later, the groups consisted respectively of 17, 10 and 11 subjects. Although the attrition rate was quite substantial and the population quite limited and numerically uneven, Babcock was able to assess the change of a number of cognitive abilities, including verbal WM. A language history questionnaire allowed her to compare the backgrounds of participants, and she tested them on the exact same four memory tests and four executive functioning tests at the start and end of the programme. While the interpreting students did not show significantly better WM at the outset of the study, they did show a significant enhancement of their verbal WM skills, contrary to the two control groups. This suggests an effect of SI training or practice on WM. Structural brain imaging showed that the SI trainees’ brains also developed differently compared to those of the other groups over the period of the study, underlying the existence of SI-training specific cerebral and behavioural changes.

Babcock's results thus point to the absence of an innate difference between interpreters and non-interpreters on WM and executive functioning skills; the differences observed being rather due to training in SI.

Babcock hypothesises that WM is likely to be used during the process of SI to "store content and rehearse pre-output translations" (2015, p. 44). Thus, according to the researcher, SI "may constitute targeted memory training which leaves the individual with increased memory span" (2015, p. 74). Babcock concludes by answering her initial question: interpreters are very likely to be unique bilinguals, with their own set of abilities. She calls for further studies on the fingerprint of the impact of SI on the brain.
3. RESEARCH

In order to further understand the relationship between WM and SI and add to the growing but scarce literature on the subject, we have decided to monitor the development of WM in students of interpreting at the Faculty of Translation and Interpretation of the University of Geneva in a longitudinal study. By comparing their WM results to that of translation students, we have been able to isolate the SI factor and see whether and how SI training and practice can have an impact on the subjects' WM.

Testing Working Memory is not an easy task, as Baddeley and Hitch's model is "inherently non-unitary in nature" (Baddeley & Logie, 1999, p. 30). As we have outlined above, according to Baddeley's model, WM is composed of a coordination unit - the central executive – and of two slave systems – the phonological loop and the visuospatial sketchpad. No one task can be said to tap into WM as a whole, and we have thus decided to select tasks which would tap into each of these three units – central executive, phonological loop and visuospatial sketchpad.

3.1 Question

We will conduct experiments which will allow us to accept or reject the null hypothesis:

Training in Simultaneous Interpretation has no significant impact on Working Memory, compared to training in Translation.

3.1a Faculty of Translation and Interpreting at the University of Geneva

The Masters in conference interpreting at the University of Geneva forms part of a long history of interpreter training at the institution. It is a founding member of the European Masters in Conference Interpreting (EMCI) consortium. This programme aims to standardise quality and entrance procedures across EMCI universities, in order to create an industry standard.

The director of the interpreting department reports that students enrolled on the masters in conference interpreting follow a three semester course, with approximately 1200 hours of interpreting time (classes and independent practice) (Seeber, May 13th, 2016). Students are taught firstly how to interpret consecutively (an interpreting technique usually combined with
note taking, a speech is first delivered, with interpreting taking place after the speaker has finished). In the second semester, students begin interpreting simultaneously (a technique used to interpret whilst the speech is being given, essentially concurrently). Assessment of students occurs continuously, with final examinations being administered after the third semester. (Hervais-Adelman et al., 2015)

3.2 Method

3.2a Participants

In order to evaluate the impact of SI training on the WM, we decided to longitudinally test two groups of masters students enrolled at the Faculty of Translation and Interpreting of the University of Geneva. The experimental group was made up of 15 interpreting students (see Table 1). They had all passed selective written and oral entrance examinations to enrol on the three-semester program.

The control group was composed of 14 translation students (see Table 1). They had all passed selective written entrance examinations to enrol on the four-semester program.

Table 1: Characteristics of both test groups.

<table>
<thead>
<tr>
<th></th>
<th>Interpreters (N=15)</th>
<th>Control (N=14)</th>
<th>Between Groups</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age in years</td>
<td>28.87 (6.37)</td>
<td>25.86 (3.12)</td>
<td>n.s</td>
<td></td>
</tr>
<tr>
<td>Range (age)</td>
<td>24.8</td>
<td>9.54</td>
<td>n.s</td>
<td></td>
</tr>
<tr>
<td>RH:LH</td>
<td>12:3</td>
<td>12:2</td>
<td>n.s</td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>18.20 (2.13)</td>
<td>18.50 (2.39)</td>
<td>n.s</td>
<td></td>
</tr>
<tr>
<td>&quot;A&quot; Languages</td>
<td>DE (2), EN (1), ES (2), FR (6), IT(2), RU(2)</td>
<td>AR(1), EN(2), ES(2), FR(6), IT(3)</td>
<td>n.s</td>
<td></td>
</tr>
<tr>
<td>Leap-Q Score</td>
<td>31.78 (5.55)</td>
<td>30.64 (6.89)</td>
<td>p=.048</td>
<td></td>
</tr>
</tbody>
</table>

( ) = Standard deviations

(A language: language towards which the translator/interpreter works, and masters as a native language)

The groups did not differ in terms of age or gender, but there was a significant difference in Leap-Q scores, which could unfortunately not be controlled for, as the participants were
recruited on a voluntary basis, a larger sample size in future study would allow for Leap-Q scores to be controlled for.

3.2b Procedure

The first task we have chosen is a verbal memory span task in which subjects are presented with sequences of units which they have to retain and repeat. As our test population possess various “A languages” (generally the subject’s mother tongue, but always a language that is mastered to native-level) (see Appendix: A), we have chosen to use a digit span task over a word recall task, in order eliminate potential linguistic disadvantage. Indeed, word-based memory tasks tap not only into WM but also into LTM when the words are known to the subject (A D Baddeley & Logie, 1999). Testing all our subjects on a verbal recall task would mean that some of them, depending on the language chosen for the task, could only rely on their WM, while others would access their LTM as well. The digit span (D-span) is thought to tap directly into the phonological loop as the subject is presented with sequences of digits, and stores them into their phonological store before retrieving it using a subvocal rehearsal system, which forms part of the phonological loop itself.

The second task we have chosen is an operation span task (O-span), in which the subject has to solve simple arithmetic problems while retaining letters being presented before repeating the sequence of letters. Mental arithmetic is a skill thought to tap into different components of WM (Ashcraft, 1995; Logie, Gilhooly, & Wynn, 1994). More precisely, mental arithmetic requires temporarily storing information (parts of results, through the phonological loop) as well as retrieving familiar calculation results and heuristic methods from LTM, thus tapping into the central executive, responsible for activating and retrieving information from LTM (Baddeley & Logie, 1999). As a result, through the O-span, we are directly tapping into our subjects’ phonological loop and central executive functions.

The third task we chose is a spatial memory task, namely the dual N-back test, in which two types of stimuli (squares and letters) are simultaneously presented on the screen and in which the subject has to remember the sequence of their on-screen position (in the case of squares) or identity (in the case of letters). They must then identify whether the current trial matches that present “N” trials previously. When N=1, this means recalling the immediately-preceding trial, when N=2, two trials previously, and so on. This task is thought to tap into the visuospatial sketchpad of WM, specifically in both its sub-components in the case of squares: the visual cache (capacity for retaining locations) at N=1, and the inner scribe (capacity for retaining sequences of movements) at N>1 (Logie, 1995)
With this selection of three tests, we are hoping to tap as much as possible into all components of WM – phonological loop, visuospatial sketchpad and central executive.

The first test session was carried out between February and April 2015, with interpreters being tested before the beginning of their training in simultaneous interpreting (SI) (they had trained exclusively in consecutive interpreting until then). All translation students were also starting their second semester.

The second test session was run in December 2015, shortly after the end of the interpreting students’ SI training, and before the start of the translation students’ fourth semester.

During the first battery of tests, we asked the participants to fill out the Language Experience and Proficiency Questionnaire (LEAP-Q), (Marian, Blumenfeld, & Kaushanskaya, 2007) which informed us about their various language proficiencies, backgrounds and language combinations (see appendix: A).

A computer suite was reserved for the tests to be carried out, the room was lit with artificial light and the blinds closed, so as to create reproducible circumstances for both the first and second tests. The computer suite used was sufficiently soundproofed as to create a quiet environment for the tests to take place. Participants were allocated a computer to use, and record of the computers used was taken so as to allow both sessions to take place upon the same machine. The participants were seated so that they could not see another active monitor, nor could they be observed by any other participants. The experimenters sat to one side silently whilst the tests took place, so as not to cause a distraction.

3.2bi Apparatus

The computers used to carry out the tests were: Dell Optiplex 9010, running Windows 7, with a monitor size of 17”. Instructions and descriptions of the test to take place were given to the participants, and verbal instructions given in English were used to clarify any further questions.
All 3 tests were programmed to run in a fixed order\(^2\) and without interruption (the participants were allowed to take breaks whilst being presented with the instructions screen for each test. The tests were programmed using The Psychology Experiment Building Language (PEBL) (Mueller, 2012) and were adapted from using the PEBL Test Battery (Mueller & Piper, 2014).

### 3.2bii D-Span

Sequences of digits, ranging from two digits (procedure outlined in Figure 2) to 12 digits, appeared on the screen at a rate of 1 every second. Participants were required to retain the digits in order and type them using the computer keyboard when prompted. If the participant successfully recalled the previous sequence, a new sequence of digits, consisting of one more element than the preceding one appeared on the screen. This incrementing of sequence length continued until the participant could no longer correctly recall the digits. After 2 incorrect trials at a certain sequence length, the test would stop. The length of the longest list of digits that the participant could correctly recall is their digit span. The digit span test took less than 5 minutes.

---

\(^2\) The order was, Test 1: D-Span, Test 2: O-Span, Test 3: Dual N-Back. We have chosen to consistently present the three tests in this order, both to participants, and in this paper.
Figure 2: D-span schematisation
3.2biii O-Span

The O-Span test requires participants to recall a sequence of letters, whilst concurrently executing a distracting arithmetical operation. The O-Span test we used is similar to that developed by Unsworth, Heitz, Schrock & Engle (2005). One notable difference was that we removed the auditory component of the test to avoid conflicts based on any linguistic expertise.

The O-Span test begins with three practice rounds, in order to familiarize the participants with the procedure and expectations. The first round demonstrates the letter recall and input phase, participants are presented with the word “ready” in the centre of the screen, which flashes 3 times (2x500ms and finally 1000ms) to centre the participants' attention on the screen. They are next presented with a sequence of letters, each letter appearing individually and staying on the screen for 1000ms. The length of the sequence in the practice rounds are 2x2 and 2x3 letters long. The participants are required to remember the letters until the end of the sequence at which point they were presented with an input grid (The procedure is outlined in figure 3).

Participants are cued to click on the letters in from the preceding sequence, in the correct order. Forgotten letters can be indicated by using the “blank” button.

The second stage of the practice introduces the distractor element to the participants. The participants are instructed that they will be shown an equation in the middle of the screen in the form (5+3 = ?). The participants are told to calculate the answer as quickly as possible before clicking “next”. Upon clicking next the participants are shown an answer to the equation, which may or may not be correct, and two buttons to select “true” and “false”. The ratio of correct and incorrect answers in the 15 practice trials was 8:7.

In this second practice round, the participants' calculation times (before clicking “next”) are measured. The mean time (± 1.25 Standard Deviation) was used as the time limit for calculations in the real experiment. Participants were unaware this was the case before the practice round, but were informed that they would have time limits in the third practice round, as well as in the real experiment.

The third practice round mimics the real test combining both the letter recall and arithmetic. The time limits derived from the second practice round were imposed, and if exceeded, the test would automatically proceed to the next letter, recording that calculation as an error.
Feedback was provided in all sections of the test, indicating both success rate in the maths and letter portions of the test. A threshold of 85% correct was imposed on the arithmetic trials, failure to maintain this level resulted in termination of the test.

The test consists of 3 sets of lengths between 3 and 7 letter/equation trials, presented in an order randomised for sequence length.
O-Span Example 1.

2 equations and 2 letter sequence.

Sequences range from 2 - 7 letters and equations and follow the pattern of:

4 + 5 = ?
9
Click True or False
W
2 + 7 = ?
11
Click True or False
X

Participants should click on the letters using the mouse in the order they were presented. The "blank" button serves to allow the sequence to be preserved whilst leaving out a forgotten letter.

You got 2/2 equations correct and 2/2 letters correct.

Figure 3: O-span schematisation
3.2biv Dual N-Back

In this task, participants observed a 3x3 grid in the centre of the screen, and attended to two concurrent series of stimuli, one comprising white squares and the other of Latin-alphabet consonants (C,H,K,N,R,W,X,Y), presented for 1s with an interstimulus interval of 3s. Squares could appear in any one of the 8 non-central positions of the grid, and consonants always appeared in the central position. Participants were required to respond whenever a stimulus from either series matched the one which appeared n steps back in the series. For both series, the value of n was the same, and was clearly indicated at the top of the screen. In the case of the squares the participants were to respond to squares appearing at the same location, and in the case of the letters they were to respond if the identity of the letters matched. The value of n ranged from 1 to 4.

For each value of n, there were 4 letter targets, 4 square targets, and 2 letter + square “dual-hit” targets, presented randomly intermixed with 10+n non-matching filler trials. The participant responded using a Swiss-standard (QWERTZ) keyboard, pressing on the right hand CTRL key for square targets, and the left hand CTRL key for letter targets, and nothing for non-targets, a “dual-hit” of both letter and square required both CTRL keys to be pressed.

The test comprised a practice round for each stimulus category separately in order to ensure that the subject had correctly understood the task. There then followed 4 test sequences with n increasing by 1 every time.

Figure 4 shows the dual N-back test.
Example at $N = 1$

Figure 4: Dual N-back schematisation
4. RESULTS

4.1 D-Span

In order to analyse the results of the D-Span test, we carried out a mixed two-by-two repeated-measures analysis of variance (repeated measures-ANOVA) including a within-subjects factor of time (two levels: first and second test) and a between-subjects factor of group (two levels: experimental and control).

The groups differed slightly on LEXP (controls: 32.94, interpreters: 37.18, unpaired t(65)=2.01, p=.048) between the groups.

![Figure 5](image.png)

We found a significant main effect of time ($F_{(1,22)}=5.205$, $p=.033$, partial eta-squared=.191), indicating that the two groups performed better at the second session. There was no significant time by group interaction ($F_{(1,22)}=2.179$, $p=.154$, partial eta-squared=.191), suggesting that there was not a differential improvement in performance as a function of group. We observe a trend indicating that the experimental group slightly outperformed the
control group during both sessions, although there was no significant main effect of group (F(1,22)=3.192, p=.088, partial eta-squared=.127).

4.2 O-Span

In order to analyse the results of the O-Span test, we carried out a mixed two-by-two repeated-measures analysis of variance (repeated measures-ANOVA) including a within-subjects factor of time (two levels: first and second test) and a between-subjects factor of group (two levels: experimental and control), for each measure.

The O-Span (or Operation Span) data contains two measures, the cumulative score of the correct memory answers (the correctly remembered letters) “OSPAN” which forms the Operation Span and the cumulative score of the mathematical element of the test “CUMUCORR”.

In the first round of testing the experimental group scored a mean O-span of 43.8 (SD: 17.03) and a mean Cumucorr of 60.9 (SD: 11.46). The control group scored an O-span mean of 37.0 (SD: 20.4) and a Cumucorr mean of 52.5 (SD:14.1). (see Figure 6 below)

In the second round of testing the experimental group scored a mean O-span of 48.6 (SD: 12.32) and a mean Cumucorr of 64.5 (SD: 6.55). The control group scored an O-span mean of 35.37 (SD: 19.74) and a Cumucorr mean of 55.75 (SD: 11.74). (see Figure 6 below)
Figure 6: O-span and Cumucorr means and standard deviations for both sessions
We found no significant effect of time, \( F_{(1,22)}=2.741, \ p=.97 \) Partial Eta Squared .268) indicating that the performance for both groups did not change. Nor was significant effect found between groups \( F_{(1,22)}=2.316, \ p=.133 \) Partial Eta Squared .236).

### 4.3 Dual N-Back

In order to analyse the results on the letter dimension, we carried out a mixed two-by-two-by-four analysis of variance (repeated measure-ANOVA), including a within-subjects factor of time (two levels: first and second test), a within-subjects factor of N (four levels, 1-4) and a between-subjects factor of group (two levels: experimental and control).

Having studied the results, it appears that participants scored very poorly on the second dimension (squares) of the N-Back test (see Appendix: B). The potential causes will be discussed in the following section of this paper. Because of the difficulties in interpreting results that might be subject to a misleading “floor-effect”, we have decided to concentrate our analysis of the N-back test on dimension 1 (letters) only.

The results of the N-back are as follows (see Figure 7):

In the first round of testing, with an “N” of 1 the experimental group scored a mean of 2.6510 (SD: 1.49), meanwhile the control group scored a mean of 2.3317 (SD: 1.23). In the second round, the experimental group scored a mean of 3.8502 (SD: 1.75) whilst the control group scored a mean of 2.8849 (SD: 0.98).

In the first round of testing, with an “N” of 2 the experimental group scored a mean of 1.0750 (SD: 1.01), meanwhile the control group scored a mean of 1.9421 (SD: 1.25). In the second round, the experimental group scored a mean of 1.55 (SD: 0.94) whilst the control group scored a mean of 1.7377 (SD: 1.31).

In the first round of testing, with an “N” of 3 the experimental group scored a mean of 1.3063 (SD: 1.13), meanwhile the control group scored a mean of 1.636 (SD: 1.66). In the second round, the experimental group scored a mean of 1.2768 (SD: 0.77) whilst the control group scored a mean of 0.9499 (SD: 1.02).

In the first round of testing, with an “N” of 4 the experimental group scored a mean of 1.1311 (SD: 1.22), meanwhile the control group scored a mean of 1.0928 (SD: 0.93). In the second
round, the experimental group scored a mean of 0.6451 (SD: 1.31) whilst the control group scored a mean of 1.36 (SD: 1.57).

(see Figure 7)

We found no significant main effect of time ($F_{(1,22)}=0.578$, $p=.455$, partial eta-squared=.026), indicating that the groups did not significantly change their performance over time. We found no significant main effect of group ($F_{(1,22)}=0.019$, $p=.893$, partial eta-squared=.001).

We notice that whilst both groups slightly improved their performance when being tested at an N of 1, the improvement in the experimental group seems to be of a greater magnitude. This improvement, however does not form part of a trend as can be seen in the graphs above.
5. Discussion and Conclusion

At the outset of this paper, we decided to investigate the nature of the relationship of the WM advantages generally found in interpreters’ WM, based on the literature referenced in our introduction. In order to delve into this complex matter, we formulated the following question: does SI training have an impact on WM?

The best way to investigate this question seemed to devise a longitudinal study comprising tests that tapped into WM, and compare the results of our experimental group (interpreting students) to that of a well-matched control group (translation students). We took advantage of our status of students of the Faculty of Translation and Interpreting of the University of Geneva, and decided to make this our field of research.

In the previous sections, we have looked at the existing literature on the topic of interpreting and WM evolution. The papers discussed in the literature review strongly indicated that professional interpreters seemed to perform better than average on WM and short-term-memory tests. Then, we specified our research question, which aimed at understanding the nature of that relationship. Do interpreters acquire a WM advantage during the course of their practice, or are students gifted with higher than average WM naturally attracted to this field? We then elaborated on the design and method of our experiments which involved two groups – an experimental and a control one - tested over time. Finally, we produced the main statistical results found for each of the three tests.

5.1. Test Results

5.1.1 D-span

As was said in the previous section, although we do find a significant effect of time, indicating that both groups significantly improved on the D-span measure, we found no significant time per group interaction. This means that both groups improved similarly in their capacity to retain sequences of numbers, and that no effect can be attributed to learning simultaneous interpreting. However, we do notice a trend: during both rounds, the experimental group slightly outperformed the control group, suggesting that potential future interpreters present a higher D-span than average, before starting to learn simultaneous interpreting.

As we explained earlier, the D-span task taps mainly into the phonological loop, a component of WM. The phonological loop is considered crucial in language comprehension, and particularly in spoken language comprehension (Clark & Clark, 1977). It would not surprising then, that students in both interpreting and translation – who both rely heavily on
language comprehension – develop their digit span through their studies. This would explain the significant effect of time we found. As for the fact that our experimental group slightly outperformed our control group on both rounds (even though the difference was not significant), we could hypothesize that since the phonological loop is thought to be even more crucial in the case of spoken versus written language comprehension, interpreting students have a slighter higher digit span than translation students due to their practice of spoken language comprehension.

5.1.2 O-span

As was said in the previous section, we found no significant effect of time and no significant effect of group. This means that the groups did not differ significantly in their results, and that they did not evolve over time. Interestingly, we notice the same trend as for the D-span: the experimental group slightly outperformed the control group. The O-span task taps into the phonological loop as well as into the central executive. It could be argued, again, that both groups rely on their phonological loop for language comprehension, with a slightly higher dependency on the phonological loop for interpreting students who deal with spoken versus written language comprehension.

5.1.3 Dual N-back

As we mentioned earlier, we have chosen to eliminate the second dimension (position of the squares) from our analysis, because the results obtained were abnormally low for both groups. We do not know exactly why this happened, but the dual N-back is subjectively a very challenging task. It could be argued that participants found themselves overloaded with input and decided to focus on one dimension, the letters. If such were the case, it would be interesting to wonder why our subjects predominantly decided to focus on the letters. Maybe they were using their phonological loop and mentally repeating phonetically the letters appearing on the screen, just as they must have done for the digit span. In any case, in further studies it would be interesting to test again translation and interpreting students on dual N-back to try and understand why we observed such results. It is important, therefore, to note that conclusions based on the N-back task in the present work apply only to the phonological, and not spatial dimension of WM.

Of course, these results could also be due to a technical problem – it is possible that the responses due to be recorded by hits on the right CTRL key (which subjects had to make in order to indicate a repetition of squares) were not properly logged.

Accordingly, we decided to focus our analysis only on the first dimension (letters). As shown in the results section, we found no significant effect of time, and no significant effect of
group. This indicates that no group performed better than the other, and that the two groups did not evolve over time. The only noticeable trend is that the experimental group slightly outperformed the control group on N=1 on both rounds, and that the experimental group improved more overtime on N=1 than the control group. However, it is not the case with different values of N.

As explained earlier, the dual N-back task taps into the visuospatial sketchpad, which is composed of two sub-components: the visual cache (for image retention, used at N=1), and the inner scribe (for sequences of movements, used at N>1). Although we found no significant effect of time nor group, we did find a positive trend for the experimental group at N=1, indicating that this group might present a better visual cache than the control group. Interestingly, they do not present a better inner scribe, which is thought to be a sub-component required for preparation for physical actions (Logie, 1995), a skill that does not seem crucial to interpreters. We are of the opinion that the N-back test when N=1 taps into a different skill than the N-back test when N>1, and that it could explain the trend observed.

Moreover, according to Brooks (1967), reading may require some form of visuospatial framework, and thus help improve one's visuospatial skills. Haenggi, Kintsch, & Gernsbacher, (1995) also argued that visuospatial capacity is needed when forming spatial mental models from written material. It is very likely that translation students, who work on written material, spend more time on average reading than interpreting students, thus also improving their visuospatial sketchpad capacity. If such is the case, it might be preferable in future studies to choose a control group that does not engage in activities likely to tap into the visuospatial sketchpad.

Overall in our 3-test experiment, we see that our experimental group did not evolve differently over time than our control group. Since the three tests chosen are accepted to be tapping into WM, this means that WM performance of the two groups did not evolve differently overtime. Training in Simultaneous interpreting was the main experimental manipulation in our research, and since it appears not to have had significant impact on the WM of participants as shown by the test we carried out, we accept the null hypothesis that SI training has no significant impact on WM

This is of course contradictory to both our initial hypotheses, and other studies, including Babcock (2015). In the following section, we will describe what the limitations of our study were, and why there is a need for further research.
5.2 Limitations and future recommendations

The current paper describes one of only two studies examining a Simultaneous Interpretation training programme and the potential link with changes to WM capacity. It adds to the understanding of the evolution of WM, but does suffer from some limitations. Like many studies involving interpreters the subject sample size was limited, due in part to the withdrawal of some trainees from the programme, but a small starting group meant this was inevitable. This could have been mitigated by performing studies across a number of intakes, allowing the results of several cohorts of interpreting students to be collated. A larger sample could also have been achieved by collaborating with other EMCI universities, ensuring that the courses followed by the students be similar. Whilst inevitable differences in education could potentially lead to different results per institution, obtaining a much larger sample size would allow for a general picture to be drawn. For the present study unfortunately neither of these two options were feasible. Future research should attempt to study larger populations.

The premise of this paper’s research question is based on the near total consistency of interpreters showing a WM advantage when compared to other groups. However, one notable difference in this study is the sampling of student interpreters at the beginning of their training; our question aiming to answer the role of training in the development of such WM. Some of the subjects tested may not ultimately become working conference interpreters, indicating that the training undertaken confer WM no advantage per se, but that it is both training and ultimately regular practice that allow for its development. It is also possible that the expected changes in WM capacity take longer to show than the time available for study. Future studies could conduct a similar, but longer-term, longitudinal study incorporating multiple sampling time points, extending into the early careers of the interpreters, monitoring changes in WM performance. The continued testing of those students who ultimately do not start regular interpreting work would also improve understanding of the advantages conferred exclusively through training, as well as a combination of training and practice.

Despite the limitations outlined above, or study lays the ground-work for future investigations of the impact of SI training on WM capacity. Since we did not find that the previously-published WM differences between interpreters and other multilingual groups can be explained by pre-training differences, we hope that future investigators will follow our lead in conducting longitudinal studies that can examine the causal relationships between SI training, expertise and WM functions.
6. Bibliography


Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and


