



## **Differences in Attainment and Performance in a Foreign Language: The Role of Working Memory Capacity**

ROGER GILABERT  
CARMEN MUÑOZ\*  
*University of Barcelona*

Received: 18 June 2010 / Accepted: 1 September 2010

### **ABSTRACT**

The goal of this study is to investigate the role of working memory capacity in L2 attainment and performance. The study uses an L1 reading span task to measure working memory of a group of 59 high-intermediate/advanced learners of English, and a film retelling task to measure their oral production. The analysis first showed a moderate to high correlation between proficiency measured by a general proficiency test and learners' fluency, lexical complexity, and accuracy but not structural complexity on the retelling task. Secondly, no correlation was found between overall proficiency and working memory. Thirdly, a weak correlation was found between fluency and lexical complexity, and working memory. When the group was split into top and bottom levels of proficiency, moderate correlations were found between lexical complexity and working memory only for the high-proficiency group. The results are discussed in the light of previous research.

### **KEY WORDS:**

Working memory, language learning aptitude, fluency, lexical complexity, structural complexity, accuracy, narrative performance.

### **RESUMEN**

El objetivo de este estudio es investigar el rol de la capacidad de memoria operativa en la proficiencia y la producción en una L2. El estudio utiliza una tarea de *reading span* en la L1 para medir la memoria operativa de un grupo de 59 estudiantes de inglés de nivel intermedio alto/avanzado, y una tarea narrativa para medir su producción oral. Los análisis muestran correlaciones significativas entre la proficiencia medida por un test de proficiencia general y la fluidez, complejidad léxica, y corrección, aunque no con la complejidad estructural. Las correlaciones también son positivas y significativas entre la memoria operativa y la fluidez y complejidad léxica, pero no se observa una correlación significativa entre la proficiencia general y la memoria operativa. Cuando se divide el grupo entre los niveles más altos y más bajos se encuentran correlaciones moderadas entre la complejidad léxica y la memoria operativa sólo para el grupo de proficiencia alta. Los resultados se analizan en base a los escasos trabajos previos.

### **PALABRAS CLAVE:**

Memoria operativa, aptitud lingüística, fluidez, complejidad léxica, complejidad estructural, corrección, producción narrativa

---

\**Address for correspondence:* University of Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona.  
E-mail: [rogergilabert@ub.edu](mailto:rogergilabert@ub.edu), [munoz@ub.edu](mailto:munoz@ub.edu)

## I. INTRODUCTION

The research reported in this paper has been conducted in the context of a larger research project (BAFIA) that aims at examining the effects of starting age of learning, input and aptitude on long-term second language learning outcomes. The BAFIA research project is a partial continuation of the BAF (Barcelona Age Factor) project (see the papers that present the final results from BAF in Muñoz, 2006). As the BAF results showed, learners' cognitive abilities play a very strong role in second language (L2) learning, and may partly at least explain the higher learning rate of older children over younger children. In particular, working memory capacity has been observed to increase monotonically between 6 and 18 years of age (Siegel, 1994) and has been claimed to be the strongest predictor of L2 learning in traditional, grammar-oriented L2 lessons (Ando et al., 1992, cited in Miyake & Friedman, 1998). More specifically, in relation to age effects in L2 learning Miyake and Friedman (1998: 347) argue that working memory is associated with faster L2 learning by older children.

The goal of this paper is to investigate the role of working memory capacity in L2 attainment and performance. The following sections introduce the area of language aptitude, focusing on memory and working memory in particular. Then the central part of the paper presents a study that used a reading span task to measure working memory of a group of high-intermediate/advanced learners of English. The discussion analyzes the results in the light of L2 performance and production theories as well as in relation to findings by previous studies on the impact of working memory capacity on L2 performance and development.

## II. BACKGROUND

### II.1 Language learning aptitude and memory

The study of language aptitude in second language acquisition (SLA) is closely linked to the use of the Modern Language Aptitude Test (MLAT), developed in the 1950s by Carroll and Sapon (1959) as a predictive measure of success in classroom L2 learning. Two decades later, and on the basis of the many studies of language aptitude conducted in that period, Carroll (1981) concluded that language aptitude is a good predictor of the rate of L2 learning. Further, Carroll characterized language aptitude as a stable learner characteristic that is composed of several relatively independent cognitive abilities, and that is not the same as verbal intelligence.

Although there exists in fact some evidence for the stability of language aptitude (Politzer & Weiss, 1969; Skehan, 1988), the view has also been defended that some prior L2

learning experience may enhance language learning aptitude (Eisenstein, 1980; Mayberry, 1993; McLaughlin, 1990). Research on this issue has been scarce and has not provided conclusive evidence either for or against stability in language aptitude (see Harley & Hart, 1997; Kormos & Sáfár, 2008; Sáfár & Kormos, 2008). In sum, although the issue is certainly relevant for the field of language teaching, there is an urgent need for more research that can show that aptitude, as McLaughlin (1990: 172) argued, has “a teachable component”.

The componential view of language aptitude is intrinsic in its traditional conceptualization. Carroll (1981) identified four components of language aptitude by means of factor analyses applied to a large number of tests that had proved to discriminate between successful and unsuccessful language learners (Carroll & Sapon, 1959). These components are: phonetic coding ability, grammatical sensitivity, inductive language learning, and rote learning ability.<sup>1</sup> More recently, Robinson (2005) has proposed that aptitude should be conceptualized as a complex and dynamic construct. In his view, aptitude is comprised of cognitive abilities that combine into higher order aptitude factors that are directly involved in various L2 learning tasks. In turn, these can be grouped into aptitude complexes that play different roles under different conditions of instructional exposure (incidental, implicit, and explicit learning). Robinson also draws attention to the need of examining the interaction between task demands, language aptitude and language learning, suggesting that different cognitive abilities are required by the information processing demands of different tasks.<sup>2</sup>

Another issue on which there exist varying perspectives is whether traditional aptitude batteries may predict or not language learning at advanced stages. Carroll (1990) and also Robinson (2005) suggest that traditional aptitude tests were not aimed at predicting very high levels of attainment. This was confirmed in the study by Winke (2005, cited in Hummel, 2009), which found a significant relationship between the MLAT subtests and L2 performance in English-speaking beginning learners of Mandarin but not in advanced learners. In contrast, Harley and Hart (1997) found significant associations between different aptitude components and L2 performance in both advanced, early immersion students and intermediate, late immersion students. In particular, their study showed a positive relationship between L2 outcomes and memory ability in early immersion students and between L2 outcomes and a measure of analytical language ability in late immersion students, which these authors interpret as reflecting age-related differences in learning or, alternatively, differences in teaching methods favored with younger and older learners respectively.

The memory test used in the MLAT is a paired-associates test that measures rote learning of foreign language material, specifically the capacity to make associations between first language (L1) lexis and target language items. This test reflects the views which prevailed in psychology at the time (Skehan, 2002), but in the light of recent cognitive

research in memory, also Carroll (1990) articulates his doubts about the validity of the associative or rote memory subtest in the MLAT.

Recently, the role of working memory has been appraised in SLA, which has led R. Ellis (2004: 534) to remark that the key issue is whether working memory is to be considered a separate individual difference factor. Some authors have suggested that working memory is a possible aptitude component (Dörnyei, 2005; Miyake & Friedman, 1998; Sawyer & Ranta, 2001). For example, Miyake and Friedman (1998: 361) argue that working memory “may be one important component of so-called language aptitude, which has repeatedly been identified in the psychometric literature, but has so far eluded precise theoretical characterization.” Furthermore, they believe that working memory can “capture the essence of the three important components of the language aptitude suggested by Skehan (1989) –a language analytic capacity, memory ability, and phonetic coding ability.” Robinson (2002: 215) sees working memory as an innovative valuable measure of aptitude, necessary for certain learning conditions (i.e. communicative classrooms) in order to “reflect the processing demands of simultaneous attention to form and meaning.” Hummel (2009), among others (Dörnyei & Skehan, 2003; N. Ellis, 1996; Gathercole & Thorn, 1998; Robinson, 2005) suggests that working memory, and more specifically the phonological memory component may be “a potentially fruitful aspect of aptitude that is not directly assessed in traditional aptitude test batteries”. However, the relationship between working memory and other aptitude components has been very sparsely researched. Robinson (2002) found that working memory as measured with a reading span test had a moderately strong correlation with language aptitude scores. In contrast, Hummel (2009) provides evidence of lack of correlations between an aspect of working memory (specifically phonological short term memory) and the aptitude components measured in the MLAT by means of the three subtests (paired associates, spelling clues, and words in sentences). Hummel underscores the distinct nature of skills underlying these measures. However, this researcher also notes that the short form of the MLAT that was used in the study did not include the number learning and phonetic script subtests, and that the latter may, to some extent, draw on phonological memory. This is clearly an issue in need of further investigation.

## **II.2 The architecture of working memory**

There have been many and varied attempts at defining the concept, architecture, and functions of working memory (for a review see Miyake & Shah, 1999), which we define here as the *mental space* where cognitive processes occur in a coordinated manner. In this article we adopt the multicomponent model of Baddeley and Hitch (1974) and also Baddeley (1984,

1986, 2000) which proposes that working memory (WM) is responsible for temporarily storing information for further processing and for subsequent integration of the outcome of those processes in the coordination and completion of goals in the performance of complex tasks. Complex tasks involving various sequences of goals include problem-solving, learning, reasoning, and language comprehension and production (McLaughlin, 1987, 1998). In the case of comprehension, for example, it has been shown that working memory capacity predicts performance in reading comprehension where, in order to comprehend a text, readers need to hold several previous propositions in memory while processing a subsequent sentence (Daneman & Carpenter, 1980). It is also well established that working memory capacity is limited and therefore differences in working memory capacity determine performance in complex tasks.

As for its architecture, in Baddeley and Hitch's (1974) and Baddeley's (1984, 1986, 2000) model working memory comprises a limited-resource central executive system which feeds on and interacts with three 'slave' passive temporary stores: the phonological loop, the visuo-spatial sketchpad, and the episodic buffer (See Figure 1 below).

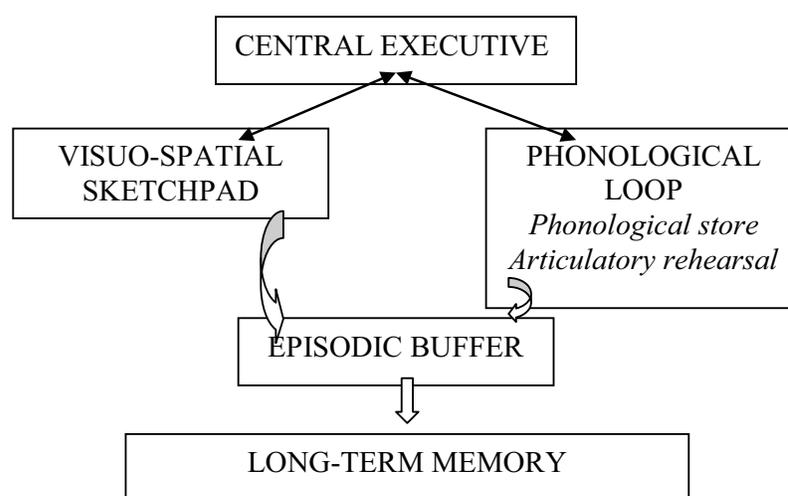


Figure 1. Model of working memory based on Baddeley and Hitch's (1974) and Baddeley's (1984, 1986, 2000).

The central executive is an attention and control system which is crucial for the coordination of the slave systems, as well as for the switching of attentional focus, the manipulation of information, directing retrieval of information from long-term memory, and for the suppression of information. The phonological loop is responsible for verbal material and it contains a phonological store which can hold information for very brief periods of time (of around two seconds) and an articulatory rehearsal component which can be used for keeping the information active through subvocal articulation in the phonological store. The

visuo-spatial sketchpad is in charge of storing and manipulating visuo-spatial information. More recently incorporated into Baddeley's model is the episodic buffer, which has been defined as responsible for integrating the material in the slave systems with information in long term memory (Baddeley, 2000). The central executive controls all cognitive activities and it also shares part of its limited capacity with the slave systems to which it can allocate more capacity. It is still an issue, however, whether both processing and storage are independent or interdependent functions of working memory, that is, whether they feed on the same or various mental resources (Friedman & Miyake, 2004).

Since in the working memory system the phonological store is responsible for temporarily storing verbal material, considerable research efforts have been devoted to the measuring capacity limits of such store in order to correlate it with L2 development. Hence, phonological memory has been found to be a strong predictor of L2 general achievement during the early elementary years (Dufva & Voeten, 1999; French, 2006; Service, 1992; Service & Kohonen, 1995), the development of L2 vocabulary in children (Gupta et al., 2003; Masoura & Gathercole, 1999, 2005) and adults (Atkins & Baddeley, 1998; Speciale, Ellis & Bywater, 2004), the development of fluency (O'Brien, Segalowitz, Freed & Collentine, 2007), and grammatical development (French & O'Brien, 2008; Williams & Lovatt, 2003). See Baddeley, Gathercole & Papagno, 1998, and N. Ellis, 1996, for reviews.

As will be seen below, in the last few years there has been a growing interest in correlating working memory as a unitary system (i.e. including both processing and storage) with both overall and individual dimensions of L2 acquisition and use.

### **II.3 Measuring working memory**

Typically researchers have used tasks which try to tap into the limits of the phonological store such as word span, digit span, non-word span, and more recently, serial non-word recognition<sup>3</sup>. All these tasks have only measured the capacity of the phonological component of working memory and not the processing component. In order to measure the role of working memory in the performance of complex tasks such as comprehension and production, complex span tasks such as the operation span task or the reading span task have been developed. Such tasks aim at taxing the information processing component in working memory while asking subjects to remember words, digits or letters. Such tasks better represent the conditions of online comprehension and production, where several pieces of information need to be temporarily held in memory while carrying out a number of other cognitive operations. Alike comprehension, language production has been shown to be a complex process involving a series of simultaneous operations requiring conscious attention

(Levelt, 1989, 1993), and this is especially the case in L2 production (Kormos, 2006). During L2 production conceptual chunks of the conceptualizer need to be held in memory while a set of other cognitive operations take place (e.g. accessing words and organizing them in a sequence, providing their morphological and phonological form, while monitoring the whole process and maintaining fluency). It is reasonable to believe that working memory tasks, such as operation span or reading span, tapping into both the processing and storage components, should help us better understand how limitations in working memory capacity explain L2 production. Another interesting but unresolved question is whether working memory capacity is task specific (e.g. a specific WM test is correlated with a specific complex task) or a general capacity that remains the same across several tasks. As Mota suggests (2003), there seems to be evidence supporting both views (see also Fortkamp, 2000, for a full discussion).

In this paper, and in line with Daneman and Carpenter (1980), a non specific-skill stance is taken. Such position sustains that more than one working memory task (i.e. operation span task, listening span task, speaking span task, or reading span task) can tap into cognitive processing be it mathematical computation, problem-solving, speaking, reading, or listening. In this paper it is therefore assumed that the reading span task, which we describe further below, stands as a general measure of working memory capacity tapping into both processing and storage.

#### **II.4 Working memory and L2 development and production**

In recent years there has been a growing interest in exploring the relationship between individual differences in working memory capacity and students' L2 production and development (Fortkamp, 2000; Guara-Tavares, 2009; Kormos & Sáfár, 2008; Mizera, 2006; Mota, 2003; Trebits & Kormos, 2008). The review that follows mainly focuses on the relationship between working memory capacity and any or all aspects of L2 speech production or its development, and does not report on the findings related to other simple measures of phonological working memory.

Fortkamp (2000) explored the relationship between working memory and oral fluency in highly advanced Portuguese learners of English. Following Daneman (1991), she used two working memory tasks, the readings span task and the speaking span task, in both the L1 of her 16 students (Portuguese) and their L2 (English), and three oral production tasks (i.e. a speech generation task consisting of a picture description; an oral reading task in which participants were asked to read a literary text aloud; and an oral slip task aimed at eliciting spoonerisms). As for fluency measures, both subjective ratings by judges and the total number of words in the time allotted were used for the speech generation task. For the oral reading

task their productions were timed, and for the oral slip task the number of spoonerisms was calculated. Fortkamp advanced three predictions: first, she predicted that learners scoring high in the speaking span task in both English and Portuguese would generate more fluent speech, would read more fluently, and would make fewer slips of the tongue; second, she suggested that the oral reading task and the oral slip task would be associated with articulation while the speech generation task would correlate better with fluency in the speech generation task; third, she predicted that the reading span task would only correlate with the reading aloud task, but not with the generation of speech task or the oral slip task. Two important findings in Fortkamp's study are that no significant correlations were observed between scores in the speaking span task and oral fluency, reading speed or spoonerisms. The only significant correlation was found between the English speaking span and fluency. She argued that this may be due to the fact that L1 language production is qualitatively different from L2 production, and so scores in the L1 speaking span task may not be associated with L2 performance. One important problem with such a finding, however, is that it is quite possible that the scores of students in the English speaking span task were affected by their proficiency, and so the results may be informing us about the relationship between participants' proficiency at a given point and their fluency rather than about their working memory capacity in relation to their fluency. The second important finding is that scores of the reading span task, both in Portuguese and English, correlated with the reading speed results in the oral reading task, but there is no report on the reading span task results in relation to oral fluency or the production of spoonerisms.

Mota (2003) extended the work of Fortkamp to include more dimensions of production, and so her study explored the relationship between working memory and L2 fluency, structural and lexical complexity, and accuracy. The study assumed that the 13 advanced learners of English with different L1s had a similar level of proficiency, but no independent test of proficiency was used. Participants performed a speaking span test in their L2 adapted from Daneman (1991) and their scores were correlated with their fluency (measured as speech rate), their accuracy (measured as errors per 100 words), their structural complexity (by calculating the total number of dependent clauses), and weighted lexical density. Both a picture description task and a narrative were used in the experiment to elicit L2 speech. Mota found working memory as measured by the speaking span task in the L2 to be significantly but moderately correlated with learners' fluency (rate but not pausing, as predicted), negatively with accuracy (i.e. learners with a higher working memory capacity made fewer mistakes), and complexity. The correlation between working memory and lexical density was negative, against Mota's prediction. As it was suggested for Fortkamp's study, a major drawback with Mota's study is the assumption that proficiency level would not affect the

performance of the working memory test in English, since even slight differences in proficiency can potentially affect performance on the test in which learners needed to produce sentences in the L2. It is therefore not surprising that working memory scores in the L2 correlated with most measures of performance. The negative correlation between speaking span scores and lexical density were explained in terms of trade-off effects, with learners prioritizing some dimensions over others during the two speaking tasks.

Mizera (2006) used three different memory tasks to explore the relationship between working memory and L2 fluency. Forty four English learners of Spanish took a speaking span task, a math span task, and a non-word repetition task, all in their L1. Mizera included an independent measure of proficiency to gauge learners' current level of attainment in Spanish. To elicit L2 speech, Mizera used a comic strip task, a word translation test, and an imitation/grammaticality test. These three production tasks were meant to elicit learners' ability to produce fluent speech, measure their lexical access speed, and tap into their monitoring abilities. Seven measures of fluency were used, tapping into different dimensions of the construct of fluency such as speed, pausing behavior, repetitions, and morpho-syntactic accuracy. Against predictions, Mizera's results showed weak correlations between working memory capacity as measured by the speaking span task and speed and monitoring, with no other correlations existing between working memory capacity and pausing, repetitions or accuracy. These same correlations became moderate when just considering the performance of the 20 most proficient students only.

In a small-scale study with 21 participants, Trebits and Kormos (2008) looked at the correlation existing between working memory capacity and L2 fluency, complexity, and accuracy on the performance of both a simple and a complex narrative task. Learners were first grouped into a high and low working memory group and then their performances were correlated with their working memory scores. A backward digit span was used as a measurement of working memory capacity, and two narrative comic strips (i.e. the simple one with clear storyline and the complex one in which learners had to make up the storyline) were used. Lexical diversity was measured in term of D-value and *Plex-lambda*, accuracy by means of the proportion of error-free clauses, and fluency by calculating speech rate. Trebit and Kormos's results revealed an impact of working memory capacity on the performance of the most complex task, with a correlation being found between backward digit span scores and both fluency and lexical complexity on the complex version of the narrative tasks.

In a pilot study tackling the relationship between working memory, pre-task planning time and L2 performance, Guara-Tavares (2009) used a speaking span task in English (the L2 of the participants) with 25 adult intermediate learners, which was later correlated with learners fluency (measured as speech rate), complexity (measured as clauses per C-unit), and

accuracy (measured as the number of errors per 100 words). Guara-Tavares found working memory scores as measured by the speaking span task in the students' L2 to be highly correlated with learners' fluency and accuracy. Again, this study did not use an independent measurement of proficiency. Such findings are therefore not surprising if we consider the arguments we provided in relation to studies that have used the speaking span tests in the L2 of the students.

Few studies have used a longitudinal design to capture gains in proficiency in relation to working memory. Two separate but interrelated studies by Payne and Whitney (2002), and Payne and Ross (2005) investigated the effects of differences in working memory capacity on students' output in the context of computer-mediated communication (CMC). Both studies aimed at measuring whether speaking can be aided by participation in online chats in combination with face-to-face interactions, and they both measured learners short-term phonological memory by means of a non-word repetition task, and working memory capacity by means of a reading span task. Intermediate learners of Spanish in a university setting, 58 in the case of the former and 24 in the latter, participated in both classroom studies. In the Payne and Whitney's study, oral proficiency was measured by means of a subjective holistic scale developed within the study, which the authors referred to as the Oral Production Interview Scale. Results in the first study showed that gains were to be associated with participants that obtained high scores in the non-word repetition task, while reading span scores did not reveal any significant differences in proficiency gains. In the Payne and Ross's study, output measures related to words, utterances and turns were used, and it was found that it was the reading span scores that correlated with the average number of words, average number of utterances, and average number of turns.

Another study that looked at the development of proficiency was conducted by Kormos and Sáfár (2008). Their goal was to study the relationship between phonological short-term memory and working memory and the development of reading, writing, listening, speaking, and use of English. In order to do so, Kormos and Sáfár used both a non-word span to measure short-term memory capacity and a backward digit span to measure working memory, which they correlated with the scores on the *Cambridge First Certificate Exam* by 121 Hungarian beginner and intermediate secondary school students from a bilingual English-Hungarian school. Kormos and Sáfár found that phonological short-term memory scores correlated with overall proficiency, writing and use of English in the case of intermediate students, while they did not find any significant correlations in the case of beginner students. As for working memory, only beginner students took the working memory test and, among other correlations, they found a weak to moderate correlation between working memory capacity and speaking ability as measured by the speaking part of the Cambridge exam. They

explained their findings in terms of the implicit learning that intermediate learners engaged in and which was captured by the non-word span, while beginner learners had experienced more explicit learning, which in turn was captured by the backward digit span but not by the phonological short-term memory non-word span.

As can be seen from the preceding literature review, there is a growing interest in understanding the relationship between working memory capacity and L2 production and attainment. These studies, however, are still scarce and their findings are not conclusive. We would like to emphasize that two major drawbacks are the use of tests in the L2 of the subjects and the lack of independent proficiency tests. Regarding the first issue, differences in L2 proficiency may influence the scores of the subject on a working memory test which is also carried out also in their L2. As for the second issue, slight differences in proficiency may have an enormous impact on the different dimensions of L2 performance, and so assuming that in a group everyone has the same or an approximate level in the L2 is methodologically problematic.

The goal of the present study is to explore the relationship between L2 learners' working memory capacity as measured by a reading span task with their overall proficiency and with their fluency, structural and lexical complexity, and accuracy during the performance of a complex narrative task. First, the experiment reported below is meant to fill a gap in the literature by looking at L2 production and attainment of L1 Spanish/Catalan speakers in an EFL context. Second, it does so by using a task that measures WM capacity in the L1 of the subjects and by controlling for their proficiency by means of an independent proficiency test, hence avoiding some of the methodological problems mentioned above.

### **III. THE STUDY**

#### **III.1 Research questions**

The study aims at answering the two following questions:

- 1) Can differences in WM capacity explain differences in general proficiency?
- 2) Can differences in WM capacity explain differences in performance (e.g. fluency, complexity, and accuracy)?

It is an open question whether working memory capacity will have an impact on overall attainment. However, based on the literature review we predict that working memory capacity may be partially responsible for learners' performance on the complex narrative task.

### III.2 Participants

From a sample of 120 undergraduate university students with L1 Catalan/Spanish, and on the basis of their proficiency scores, data from 59 students corresponding to the top (N=31) and bottom (N=28) quartiles were used in the study. These included 50 female and 9 male students, their average age being 22.9. They had all gone through the Catalan educational system which offers English as a foreign language which, despite its more communicative approach in recent years, is still largely based on explicit grammar instruction. It was also found that there were considerable differences in their degrees of exposure to L2 input, ranging from people who had mainly been exposed to the language in just the formal context of school to people who had spent varying periods of time in English-speaking countries.

### III.3 The BAFIA battery of tests and tasks

Students in the overall BAFIA study exploring the effects of age, input, and aptitude on L2 acquisition took the following battery of tasks: a personal information and input questionnaire exploring their exposure to and their learning experience with English; three proficiency tests (a standardized general proficiency test OPT – *Oxford Placement Test* –, two vocabulary tests – X\_lex and Y\_lex (Meara & Milton, 2003), which measure vocabulary size and which have been shown to strongly correlate with proficiency (Meara & Jones, 1988; Miralpeix, 2009) –, and a phonetic categorization task; three aptitude-related tests (including a lexical access test, a non-word repetition task, and a reading span task) all developed and tested within the BAFIA project; two standard tasks to elicit L2 performance (an oral interview on their experience learning English and a film retelling task).

In this exploratory correlational study, the OPT was used as a measure of proficiency. This was correlated with the learners' scores on the reading span task. Both proficiency and working memory scores were then correlated with learners' performance on the film retelling task.

#### III.3.1 The reading span task

As mentioned before, evidence suggests that several complex memory tasks tap into working memory capacity. In the BAFIA project, we needed a complex working memory task for us to be able to correlate it with different dimensions of performance. To our knowledge, there had been two attempts to adapt the reading span task to Spanish (Sagarra, 2002), and this served as a basis for the development of a reading span task for both Spanish and Catalan. In the reading span task subjects are presented with a sentence and they are instructed to perform two tasks: first, to say whether each sentence makes sense or not and, second, to remember

the last word in each sentence. The two jobs are meant to have the learners hold words in memory as they process new incoming information. Eighty sentences in the L1 of the student (i.e. either Spanish and Catalan) were created, 50% of which made sense while the other 50% did not. Sentences ranged from 8 to 12 words and they were organized into 20 blocks with 3 trial blocks. They were presented in series of 2 to 6 sentences. Following Waters and Caplan (1996), 4 types of syntactic structures were used (e.g. *El aspirante a profesor fue quien hizo el examen* -- cleft subject: CS; *El abuelo fue lo que la canción de la infancia recordó* -- cleft object: CO; *La melodía que vino del bosque la cantaron los pájaros* -- object± subject: OS; *El libro que el lector compró hojeó las páginas.* -- subject± object: SO), which require different levels of processing. To avoid any effects of word length and word frequency, words to be recalled were selected on the basis of their similar frequency (Sebastián, et al. 2000), they were all 3-syllable words, and included no proper names nor abstract concepts. Following Friedman and Miyake (2004) and Conway et al. (2005), in order to compute the score of each subject in the reading span task every element (i.e. word) within the same unit (i.e. number of sentences ranging from 2 to 6) remembered by the subject was proportionally calculated (i.e. a word in a 6-sentence unit has a higher value than a word remembered in a two-sentence unit). Accuracy, order, and reaction times were also taken into consideration<sup>4</sup>. During piloting, internal consistency reached a Cronbach's alpha of .872.

### **III.3.2 Film retelling task**

For this study, only data from the narrative task were used. A clip of Chaplin's *Modern Times* was used to elicit participants' L2 speech. Since the scene depicted in the clip was a complex one with several characters and a considerable number of events, participants were shown the movie twice. After viewing the clip for the second time, participants were asked to retell what had happened in the story. Task performance happened in the presence of a researcher who was instructed not to interact with the subjects, who had no time limitations to complete the task. Occasional participation by the researchers was detected for backchannelling (e.g. uhm, uhm) and to provide the odd word when subjects got stuck during the narrative. These interventions were not systematic and were not thought affect participants' production.

### **III.4 L2 proficiency and performance measures**

As mentioned above, a standardized *Oxford Placement Test* was used as a measure of participants' proficiency. As for L2 performance, all narrative samples were transcribed by means of the CLAN mode (software used for the transcription and analysis of speech) of the

CHILDES database. Four different measures of performance were used each tapping into a different dimension. Fluency was calculated in terms of syllables per minute in the participants' unpruned speech. This is a standard measure in the literature which taps into both speed and pausing, and so it provides a global idea of L2 speakers' fluency (see Skehan, 2009). Lexical complexity was measured by means of Guiraud's Index, a corrected version of the type/token ratio which compensates for text length. Again this is a standardized measurement in the literature which taps into the variety of vocabulary used by L2 speakers. Structural complexity was measured as the number of sentence nodes per AS-unit, which is a stereotypical measure of subordination. Accuracy was computed by calculating the number of errors every 100 words, which is a measurement often used in psycholinguistic research. Calculating errors per 100 words also has the advantage of compensating for differences in text length. A composite measure averaging the score of the four different dimensions was also obtained. As for the coding, interrater measures were used on 10% of the data with overall interrater reliability reaching 92%.

#### IV. RESULTS

	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Minimum</b>	<b>Maximum</b>
<b>OPT</b>	59	44.64	8.042	31	57
<b>Reading Span score</b>	59	.56	.13	.22	.81
<b>Syllables x minute</b>	59	139.29	32.67	65.66	230.65
<b>Guiraud's Index</b>	59	7.33	.93	5.95	9.64
<b>SN per AS-Units</b>	59	1.74	.23	1.33	2.32
<b>Accuracy</b>	59	4.85	2.94	1	13
<b>Average performance (z scores)</b>	59	.026	.701	-1.59	1.08

OPT = *Oxford Placement Test*; SN = *Sentence Nodes*; AS-unit = *Analysis of Speech Unit*.

Table 1. Descriptive statistics of the whole group

<b>Low proficiency group</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Minimum</b>	<b>Maximum</b>
<b>OPT</b>	28	37.00	3.8	31	43
<b>Reading Span score</b>	28	.55	.13	.31	.78
<b>Syllables x minute</b>	28	124.03	27.3	65.66	181.44
<b>Guiraud's Index</b>	28	6.8	.66	5.95	8.28
<b>SN per AS-Units</b>	28	1.7	.25	1.33	2.32
<b>Accuracy</b>	28	7.07	2.3	3	13
<b>Average performance (z scores)</b>	28	-.473	.617	-1.59	.75

<b>High proficiency group</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Minimum</b>	<b>Maximum</b>
<b>OPT</b>	31	51.55	2.8	45	57
<b>Reading Span score</b>	31	.57	.14	.22	.81
<b>Syllables x minute</b>	31	152.51	31.44	98.37	230.65
<b>Guiraud's Index</b>	31	7.75	.95	5.98	9.64
<b>SN per AS-Units</b>	31	1.78	.20	1.43	2.26
<b>Accuracy</b>	31	2.99	1.93	1	11
<b>Average performance (z scores)</b>	31	.444	.450	-.65	1.08

Table 2. Descriptive statistics by proficiency level

First of all, we explored the relationship between proficiency and performance, a relationship that is more often assumed to exist than tested. Because the sample was not normally distributed, since it contained the participants in the bottom and top quartiles as mentioned above, Spearman correlations were used. As displayed in Table 3 below, when overall proficiency is correlated with each dimension of performance it can be seen that there exists a moderate correlation between overall proficiency and both fluency and lexical complexity, while there is no correlation between proficiency and structural complexity, and there is a moderately strong correlation between proficiency and accuracy as measured by the number of errors per 100 words.

	<b>Fluency</b> Unpruned speech rate	<b>Lexical</b> <b>Complexity</b> Guiraud	<b>Structural</b> <b>Complexity</b> SN x AS Units	<b>Accuracy</b> Errors x 100 words
<b>Proficiency</b>				
<b>Spearman rho</b>	<b>.456**</b>	<b>.573**</b>	.155	<b>706 **</b>
<b>Sig (2-tailed)</b>	.000	.000	.253	000
<b>N</b>	59	59	59	59

Table 3. Spearman correlations between participants' overall proficiency and fluency, lexical and structural complexity, and accuracy

Our first research question concerned the relationship between working memory as measured by the reading span task and overall proficiency. As seen in Table 4 below, no correlation was found, suggesting that the participants' current level of attainment cannot be explained in terms of working memory capacity alone.

		<b>WM</b> <b>capacity</b>	<b>Average</b> <b>Proficiency</b>
<b>WM capacity</b>	<b>Spearman rho</b>	1	.077
	<b>Sig (2-tailed)</b>		.560
	<b>N</b>	59	59
<b>Proficiency</b>	<b>Spearman rho</b>	.077	1
	<b>Sig (2-tailed)</b>	.560	
	<b>N</b>	59	59

Table 4. Spearman correlations between working memory capacity scores and proficiency

Our second research question asked whether there exists a correlation between working memory capacity and participants' L2 fluency, lexical and structural complexity, and accuracy. Table 5 below shows the correlations existing between working memory capacity scores and the different dimensions of production individually.

	<b>Fluency</b> Unpruned speech rate	<b>Lexical</b> <b>Complexity</b> Guiraud	<b>Structural</b> <b>Complexity</b> SN x AS Units	<b>Accuracy</b> Errors x 100 words
<b>WM Capacity</b>				
<b>Spearman rho</b>	<b>.231*</b>	<b>.266*</b>	.117	.042
<b>Sig (1-tailed)</b>	.043	.023	.195	.378
<b>N</b>	59	59	59	59

Table 5. Spearman correlations between working memory capacity scores and fluency, lexical complexity, structural complexity, and accuracy on the narrative task

The individual correlations between working memory capacity and the different dimensions of performance show that the two dimensions that explain the correlation between working memory capacity and overall performance are fluency and lexical variety, while no correlation exists between working memory and structural complexity or accuracy. These results suggest that participants with a higher working memory were more fluent and used a wider variety of vocabulary than learners with a lower working memory capacity.

Finally, Table 6 below details the correlations between working memory capacity and the different dimensions of production by proficiency groups. Results highlight that differences in working memory capacity explain the use of a wider variety of words during performance, with a moderate correlation existing between working memory capacity scores and lexical complexity only for the high proficiency group.

	<b>Fluency</b> Unpruned speech rate	<b>Lexical</b> <b>Complexity</b> Guiraud	<b>Structural</b> <b>Complexity</b> SN x AS Units	<b>Accuracy</b> Errors x 100 words
<b>WM Capacity Low proficiency group</b>				
<b>Spearman rho</b>	.202	-.022	-.068	-.0993
<b>Sig (1-tailed)</b>	.161	.458	.371	.326
<b>N</b>	28	28	28	28
<b>WM Capacity High proficiency group</b>				
<b>Spearman rho</b>	.186	<b>.543**</b>	.217	.100
<b>Sig (1-tailed)</b>	.162	.001	.124	.297
<b>N</b>	31	31	31	31

Table 6. Spearman correlations between working memory capacity and fluency, lexical and structural complexity, and accuracy on the narrative task separated by proficiency groups

## V. DISCUSSION

It is reasonable to believe that learners with a higher proficiency level can perform better than learners with a lower proficiency level when carrying out a complex narrative. Second language production models (Kormos, 2006) have shown us that more proficient learners resemble L1 speakers in the sense that they have a more complete system (e.g. a larger lexicon and a better specified grammatical system) than lower proficiency learners, and have partially automatized some of the processes (e.g. lexical access or grammatical and phonological encoding) that require higher attentional and memory resources. The correlations between overall proficiency and each individual dimension of performance have indicated that higher overall proficiency is to be associated with higher fluency, lexical

complexity, and accuracy, but not with structural complexity. Two explanations seem possible. First, our measurement of sentence nodes per AS units may not have been a sensitive measure to capture differences among learners with different levels of proficiency (see Norris & Ortega, 2009 for a discussion on the challenges of measuring structural complexity). Alternatively, structural complexity does not necessarily correlate with proficiency, and so more proficient learners may not necessarily use more syntactically complex structures (Pallotti, 2009).

The first research question explored the relationship between working memory and overall attainment. The results in this study suggest that there is no correlation between working memory and overall attainment, and are in line with those of Mizera (2006) who found no correlation between a general proficiency test in L2 Spanish and working memory scores. In the first place, the present study tested the proficiency of learners at a given point after years of learning the language and was therefore not looking at gains over time as in Kormos and Sáfár (2008). It is feasible to believe that differences in working memory alone, which as we saw in the literature review have been shown to affect the rate of acquisition, cannot explain the different levels of attainment, and should therefore be combined with other variables. In a study with participants in the larger project, Muñoz (in press) found that while age of onset does not seem to explain learners' progress in the long term, certain conditions of exposure to input correlate with attainment. Secondly, the results in this study confirm that learners may progress in the learning of an L2 regardless of their working memory capacity, even if the rate at which they do so may vary.

The second question concerned the relationship between working memory capacity and performance on a complex narrative task. There is no question that speaking in an L2 is a complex cognitive task and, as both L1 and L2 production models (Kormos, 2006; Levelt, 1989, 1993) have indicated, some of the processes involved in speaking require attentional and memory resources. Hence, both selection and organization (i.e. macro- and micro-planning) of concepts prior to formulation require conscious attention. Additionally, this information must be kept in working memory as formulation and articulation proceed. As for formulation, while in the L1 vocabulary access, grammatical encoding, and morpho-phonological encoding are largely automatic and therefore do not consume considerable attentional and memory resources, in the L2 formulation tends to be more laborious and often requires serial rather than automatic and parallel processing. Additionally, monitoring also takes a toll on attentional and memory resources when supervising the pre-verbal plan, the pre-articulatory plan, or overt speech. So as they produce the language L2 speakers must hold the products of those processes in memory while generating and planning new messages. Better coordination of processing and storage, due to a larger working memory capacity,

partially explains why some learners perform better than others. When we examined the different dimensions individually, it was seen that it was fluency and lexical complexity that correlated significantly with working memory capacity. Our results are in line with previous studies (Fortkamp, 2000; Mizera, 2006) which showed weak correlations between working memory and fluency. It is possible that higher working memory capacity is associated with faster lexical access and retrieval, which as a consequence have a positive effect on fluency. However, as Mizera (2006) argues, it is also possible that ‘vocabulary depth’, which is the consequence of richer semantic associations growing alongside proficiency, may explain faster lexical access and retrieval better than working memory capacity. As for lexical complexity, during L2 production learners must find and retrieve the words that better correspond to their intended message. Learners with a higher working memory capacity may not only access and retrieve words faster, but they may also have enough cognitive resources to consider several alternatives. The mild correlation between working memory capacity and lexical complexity may suggest just that. However, proficiency has shown to be a better predictor of lexical variety, with a stronger correlation than that of working memory, probably because it ensures a larger vocabulary with more and stronger interconnections. In sum, proficiency understood as a ‘wider’ and ‘deeper’ vocabulary may play a larger role than working memory capacity in explaining higher fluency and higher lexical complexity.

Finally, results of the effects of working memory capacity on performance with participants in the top and bottom proficiency quartiles revealed an interesting picture. It was seen that differences in working memory were significantly correlated with the performance of only the participants in the high proficiency group (.531) and not with the performance of participants in the lower proficiency group. As Mizera (2006:107) suggests: “Another possibility is that working memory is most helpful to L2 learners at later stages of acquisition.” At earlier stages of development, learners may need to use more cognitive resources to store intermediate products between conceptualization and formulation than advanced learners. Learners may need to be using considerable attentional and memory resources to access and retrieve words, and to give messages their syntactic and morpho-phonological shape, to a point that differences in working memory may not matter. Our results therefore suggest that for working memory to pose an advantage in L2 performance a certain level of proficiency needs to be achieved, results which are in contrast with those of Kormos and Sáfár (2008). Therefore, more research is needed that looks at the relationship between working memory and proficiency level with higher control of the latter. On the other hand, this study has confirmed previous findings concerning the association between working memory capacity and lexical complexity, since when breaking down performance into its

different dimensions, it was found that lexical complexity is the performance dimension that most clearly benefits from a higher working memory capacity.

## VI. LIMITATIONS, FUTURE RESEARCH, AND IMPLICATIONS

Some limitations to the present study can be highlighted. First, even if it is still an unresolved issue whether processing and phonological storage are independent or interrelated, future research should incorporate more than one working memory task, tapping into both of them. The choice of the reading span task in this research project may have only told part of the picture, since it looks at processing and storage simultaneously, and the future use of other tasks more specifically tapping into the processing (e.g. attention-switching task) and the storage component (e.g. non-word recognition task) may enrich our understanding of the effects of individual differences in working memory on attainment and performance. Secondly, even if this was an exploratory study, more measures of performance could be used to capture different aspects of the dimensions of production (see Housen & Kuiken, 2009 for a review of performance and developmental measures). Thirdly, it is an issue whether the amount of content conveyed by participants may have affected their performance, with for example some participants choosing to say less in order to maintain an acceptable level of performance. Future studies should try to incorporate 'task completion' measures that may be analyzed alongside and in relation to performance measures. In the fourth place, potential trade-off effects among the different dimensions of production should be investigated. The L2 performance literature (Skehan, 2009) has shown us that L2 speakers tend to prioritize some dimensions over others and this may be related to their working memory capacity. In the fifth place, longitudinal data should be used in order to provide more information about attainment. Finally, this study has not aimed at finding answers for the more general questions concerning the relationship between working memory and aptitude, or the issue of whether pedagogical interventions can improve aptitude and working memory in particular. Future research is needed both in this direction as well as in the direction of the present exploratory study.

However, we may also extract some pedagogical implications from the results obtained in this study. A pedagogical implication is the confirmation that attentional and memory resources are limited and therefore students at lower levels of proficiency may be overwhelmed by the tasks at hand. In this study, our results showed that, whatever advantages a high WM capacity may pose in the performance of a complex task, lower proficiency learners could not make use of them to better perform on a narrative film retelling. Only when their proficiency was higher, WM capacity could discriminate performance, with high

WM students using a wider range of vocabulary during narrative performance. This has implications, among others, for task selection and sequencing in syllabus design. It seems reasonable to believe that at lower stages less cognitively demanding tasks that do not tax learners' resources may contribute to smoother and more balanced performance and to more attentional and memory resources being devoted to the different dimensions of production.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the grants HUM2007-64302 and 2009SGR137, by the Spanish *Ministerio de Educación*, and the *Agència de Gestió dels Ajuts Universitaris i de Recerca de la Generalitat de Catalunya*, respectively. We also give special thanks to Mayya Levkina for her help with the reading span task and her general assistance.

---

## NOTES

<sup>1</sup> Other aptitude measurement batteries have been produced, although MLAT remains the most influential one. Also widely used, Pimsleur's Language Aptitude Battery (PLAB) (Pimsleur, 1966) was specifically designed to measure high-school students' language learning aptitude. Grigorenko, Sternberg and Ehrman (2000) developed the CANAL-FT following a new conceptualization of language learning aptitude as the ability to cope with novel experience in language learning.

<sup>2</sup> This is the approach taken in the elaboration of the High-Level Language Aptitude Battery (Hi-LAB) that is currently under way at the Center for Advanced Study of Language (CASL) at the University of Maryland (Doughty, in press).

<sup>3</sup> All these tasks are based on either the recognition or recall of single units, such as words, digits, or letters and are meant to tap into (phonological) short-term memory capacity. The digit span uses a series of digits (that may range from 2 to 9 digits) that subjects are presented with and asked to recall. The word span uses single words that increase in length while the non-word span uses non-words that subjects are asked to recall. The serial non-word recognition does not ask subjects to recall non-words but rather to recognize differences between series of non-words. Non-words may be L1-like, L2-like, or created on the basis of a language which is completely alien to the subject. In all these kinds of tasks subjects are not engaged in more complex mental operations, such as comprehending or producing language.

<sup>4</sup> This means that if subjects did not remember a word correctly, words were remembered in the wrong order, or they took longer than 5000 milliseconds to answer, they were penalized and therefore the word was not counted.

## REFERENCES

- Ando, J., Fukunaga, N. Kurahashi, J., Suto, T. Nakano, T. & Kage, M. (1992). A comparative study on the two EFL teaching methods: The communicative and the grammatical approach. *Japanese Journal of Educational Psychology*, 40, 247-256. (in Japanese)
- Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19, 537-552.
- Baddeley, A.D. (1984). The fractionation of human memory. *Psychological Medicine*, 14, 259 - 264
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4(11), 417-423.

- Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158-173.
- Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G. Bower (Ed.), *The Psychology of Learning and Motivation*. (Vol. 8) (pp. 47-89). New York: academic Press.
- Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual Differences and Universals in Language Learning Aptitude* (pp. 119-154). Rowley, MA: Newbury House.
- Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs, NJ: Prentice Hall.
- Carroll, J. B. & Sapon, S. M. (1959). *The Modern Language Aptitude Tests*. San Antonio, TX: Psychological Corporation.
- Conway, A., Kane, M., Bunting, M., Hambrick, D.Z, Wilhelm, O., & Engle, R. (2005) Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786.
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, *20*, 445-464.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*, 450-466.
- Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 589-630). Malden, MA: Blackwell.
- Doughty, C. (in press). Testing Aptitude in Foreign Language Learners: two situations and applications. Research report. <http://www.recercat.net/handle/2072/1761>
- Dufva, M., & Voeten, M. J. M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, *20*(3), 329-348.
- Eisenstein, M. (1980). Childhood bilingualism and adult language learning aptitude. *International Review of Applied Psychology*, *29*, 159-174.
- Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91-126.
- Ellis, R. (2004). Individual Differences in Second Language Learning. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 525-551). Oxford: Blackwell.
- Fortkamp, M. B. M. (2000). *Working memory capacity and L2 speech production: an exploratory study*. Unpublished PhD Dissertation, Universidade Federal de Santa Catarina, Florianópolis.
- French, L. M. (2006). *Phonological Working Memory and Second Language Acquisition in Children*. Lampeter: Edwin Mellen Press.
- French, L. M. & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*(3), 463-487.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and inference control functions: A latent variable analysis. *Journal of experimental Psychology: General*, *133*, 101-135.
- Gathercole, S. E. & Thorn, A. (1998). Phonological short-term memory and foreign language learning. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 141-158). Mahwah, N.J.: Erlbaum.
- Grigorenko, E. L., Sternberg, R. J. & Ehrman, M. E. (2000). A theory based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *Modern Language Journal*, *84*, 390-405.
- Guará-Tavares, M. G. (2009). The relationship among pre-task planning, working memory capacity, and L2 speech performance: a pilot study. *Linguagem & Ensino*, *12*(1), 165-194.
- Gupta, P., MacWhinney, B., Feldman, H. M., and Sacco, K. (2003). Phonological Memory and Vocabulary Learning in Children with Focal Lesions. *Brain and Language*, *87*, 241-252.
- Harley, B. & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*, 379-400.

- Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 461-473.
- Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, 30, 225-249.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah NJ: Lawrence Erlbaum.
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(2), 261-271.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA; MIT Press.
- Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, & C. Wallesch (eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: de Gruyter.
- Masoura, E. V., & Gathercole, S. E., (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology*, 34, 383-388.
- Masoura, E. V., & Gathercole, S. E. (2005). Phonological short-term, memory skills and new word learning in young Greek children. *Memory*, 13, 422-429.
- Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American sign language. *Journal of Speech and Hearing Research*, 36, 486-512.
- McLaughlin, B. (1987). *Theories of second language learning*. London: Edward Arnold.
- McLaughlin, B. (1990). The relationship between first and second languages: Language-proficiency and language aptitude. In B. Harley, P. Allen, J. Cummins & M. Swain (Eds.), *The Development of Second Language Proficiency* (pp. 158-178). Cambridge: Cambridge University Press.
- McLaughlin, B. (1998). Second language learning revisited: the psycholinguistic perspective. In A. F. Healy & L. E. Jr. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 399-411). Mahwah, NJ: Lawrence Erlbaum.
- Meara, P. & Jones, G., (1988). Vocabulary size as a placement indicator. In Grunwell, P. (Ed.), *Applied Linguistics in Society CILT* (pp. 80-87). London.
- Meara, P. M. & Milton, J. L. (2003). *X\_Lex. The Swansea Vocabulary Levels Test*. Newbury: Express.
- Miralpeix, I. (2009). Measuring vocabulary size: New tools for assessment. *CRAL Conference*. 2009, October 19-31, Logroño, Spain.
- Miyake, A. & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A.F. Healy & L.E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, N.J.: Erlbaum.
- Miyake, A., & Shah, P. (1999). Models of working memory: An introduction. In: A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp.1-27). Cambridge, U.K.: Cambridge University Press.
- Mizera, G. J. (2006). Working memory and L2 oral fluency. PhD Dissertation. University of Pittsburgh.
- Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos*, 24, 69-104.
- Muñoz, C. (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters Ltd.
- Muñoz, C. (in press). Input in foreign language learning: more significant than starting age? *International Review of Applied Linguistics (IRAL)*
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological Memory Predicts Second Language Oral Fluency Gains in Adults. *Studies in Second Language Acquisition*, 29, 557-581.
- Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, 30(4), 590-601.

- Payne, S.J. & Ross, B. M (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning and Technology*, 9K(3), 33-54.
- Payne, S.J. & Witney, P. (2002). Developing L2 oral proficiency through synchronous CMC: output, working memory, and interlanguage development. *CALICO Journal*, 20(1), 7-32.
- Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery*. New York: Harcourt Brace Jovanovich.
- Politzer, R. L. & Weiss, L. (1969). *An experiment in improving achievement in foreign language learning through learning of selected skills associated with language aptitude*. Stanford, CA.: Stanford University,
- Robinson, P. (2002). Individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfeld, and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins.
- Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46-73.
- Sáfár, A. & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics*, 46(2), 113-136.
- Sagarra, N. (2002). The role of syntactic modifications on L2 oral comprehension. In C. Wiltshire & J. Camps (Eds.), *Romance Phonology and Variation* (pp. 197-210). Amsterdam: John Benjamins.
- Sawyer, M. & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 319-353). Cambridge: Cambridge University Press.
- Sebastián, N., Martí, M. A., Carreiras, M. F. & F. Cuetos, F. (2000). *LEXESP: Léxico Informatizado del Español*. Ediciones de la Universidad de Barcelona, Barcelona.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45-A, 21-50.
- Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16, 155-172.
- Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development*, 17, 109-124.
- Skehan, P. (1988). *A comparison of first and foreign language learning ability: A follow-up to the Bristol Language Project* (ESOL Working Document No. 8). London: London University, Institute of Education.
- Skehan, P. (1989) *Individual Differences in Second Language Learning*. London: Edward Arnold.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 69-93). Amsterdam: John Benjamins.
- Skehan, P. (2009). Modeling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30(4), 510-532.
- Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, 25, 293-321.
- Trebits, A., & Kormos, J. (2008). Working Memory Capacity and Narrative Task Performance. *Proceedings from the 33<sup>rd</sup> International LAUD Symposium*, Landau/Pfalz, Germany.
- Waters, G. S. & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review*, 4, 761-772.
- Williams, J. N., & Lovatt, P. J. (2003). Phonological Memory and Rule Learning. *Language Learning*, 53(1), 67-121.
- Winke, P. (2005). *Individual differences in adult Chinese second language acquisition: The relationships among aptitude, memory, and strategies for learning*. Unpublished PhD dissertation, Georgetown University.