Lang. Teach. (2011), **44.2**, 137–166 © Cambridge University Press 2011
doi:10.1017/S0261444810000509

State-of-the-Art Article

Aspects of working memory in L2 learning

Alan Juffs University of Pittsburgh, Pittsburgh, USA
juffs@pitt.edu

Michael Harrington University of Queensland, Australia
m.harrington@uq.edu.au

This article reviews research on working memory (WM) and its use in second language (L2) acquisition research. Recent developments in the model and issues surrounding the operationalization of the construct itself are presented, followed by a discussion of various methods of measuring WM. These methods include word and digit span tasks, reading, listening and speaking span tasks. We next outline the role proposed for WM in explaining individual differences in L2 learning processes and outcomes, including sentence processing, reading, speaking, lexical development and general proficiency. Key findings are that WM is not a unitary construct and that its role varies depending on the age of the L2 learners, the task and the linguistic domain. Some tests of WM may in fact be tests of differences in ability to attend to aspects of the L2. Future research will focus on matching tests of WM more closely with linguistic tasks and using more standardized, replicable measures of WM in new areas including writing in non-alphabetic scripts, instructional interventions and cognitive neuropsychology.

1. Introduction

The learning and use of a second language (L2) draws on a range of cognitive processes. One process receiving increasing attention is working memory. This is the system for temporarily storing and processing information while performing higher order cognitive tasks such as comprehension, learning and reasoning (Baddeley & Logie 1999). Since first proposed by Baddeley & Hitch (1974), the construct has been the focus of extensive research in cognitive psychology and psycholinguistics. Differences in working memory capacity (WMC) have been related to performance in cognition and language in experimental, neuro-cognitive and individual differences research (Miyake & Shah 1999; Oberauer, Süß, Wilhelm & Wittman 2003). Over the past two decades WM has also attracted the attention of L2 researchers, who have been interested in how individual differences in WMC might account for variation in L2 learning and use. Researchers have examined the role of WM in various L2 language processes, including reading (Walter 2006, Leeser 2007), writing (Adams & Guillot 2008),

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

sentence processing (Juffs 2004; Felser & Roberts 2007), speaking (Fortkamp 1999; O'Brien et al. 2006), vocabulary development (Papagno & Vallar 1995; Cheung 1996), learning grammar (Williams & Lovatt 2003) and the processing of input and intake (Mackey et al. 2002). WM has also been implicated as a core element in L2 aptitude (Dörnyei & Skehan 2003; Robinson 2005) and as a predictor of overall proficiency (van den Noort, Bosch & Hugdahl 2006). There has also been research examining the interaction between WM and affective factors like extraversion (Dewaele, Petrides & Furnham 2008) and stress (Rai et al. in press). In this article we provide an overview of this literature. The aim is to identify current and emerging areas of WMC research for researchers and practitioners interested in a better understanding of the role WM plays in L2 learning and use. The article is organized as follows. In the next section (Part 2) we describe the WM construct as applied in L2 research and present the methodology used to study and measure it. In Part 3 we review research on WMC as a factor in L2 performance, first considering WM as a general constraint on L2 learning outcomes, including its proposed relationship to L2 aptitude, and then examining the effect of individual differences in WMC in specific domains of L2 learning and use. Finally, in Part 4, we will discuss future directions in L2 WM research, including possible applications in educational settings.

2. What is working memory?

Working memory refers to the mental processes responsible for the temporary storage and manipulation of information in the course of on-going processing. Working memory is not memory PER SE, if by that term we simply mean the capacity to store the products of our experience in the world. Rather, it is better understood as a system that controls and regulates behavior, or, in much more general terms, as '. . .a set of limiting factors for performance in cognitive tasks' (Oberauer et al. 2003: 168). These factors arise from the interaction of a complex web of neural and cognitive processes, leading some researchers to suggest that WM is better understood as a part of cognition rather than as a part of memory (Engle 2007).

WM involves both the storage and processing of information. Storage elements consist of the traditional short-term and long-term memory stores and will be described first. Then the processing function will be described. This element is responsible for directing attentional resources and the interplay between the storage domains and executive control. The best-known model of WM is that first proposed by Baddeley & Hitch (1974); a later version from Baddeley (2000) is presented in Figure 1.

The original model had three elements. It was comprised of two short-term storage domains consisting of the PHONOLOGICAL LOOP and the VISUO-SPATIAL SKETCHPAD, and a CENTRAL EXECUTIVE controlling the flow of information between these domains and other cognitive processes. The phonological loop handles phonological and verbal information, while the visuo-spatial sketchpad processes visual and spatial information. Later a third element, THE EPISODIC BUFFER, was added (Baddeley 2000) as the place where various types of information are temporarily stored and integrated. The three short-term storage domains are called 'slave' systems to denote their passive roles as repositories of information controlled by the central executive.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

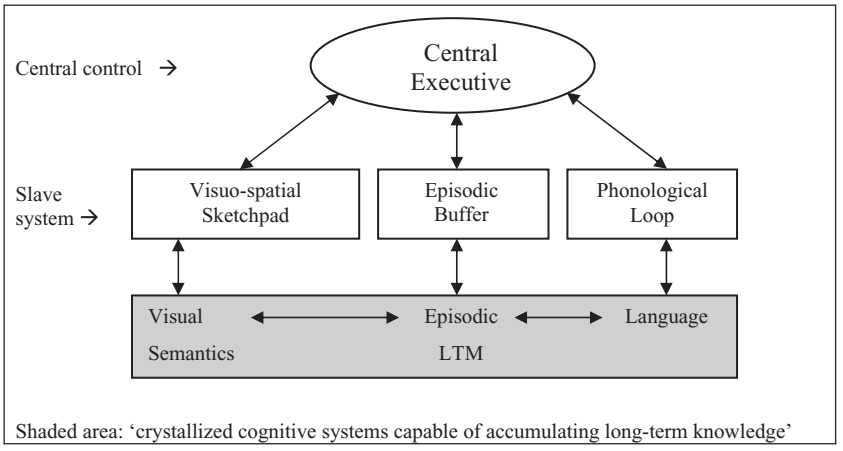


Figure 1 Development of the Working Memory Model (Baddeley 2000).

The short-term memory stores hold a limited amount of information that is available only for a matter of seconds before it is lost. This capacity is limited both by how much information can be maintained and how long that information is available. The processing of verbal material depends on the phonological loop and this element of it has, from the outset, received the greatest attention from researchers interested in language learning and processing. The shaded area in Figure 1 represents long-term storage into which elements in the WM component above may be able to penetrate or gain access. Essentially, WMC is perhaps best thought of as a bottle-neck through which information has to pass in order to be permanently stored in long-term memory.

2.1. Phonological memory

The processing of phonological information is central to language learning and use, so the phonological short-term memory, or PHONOLOGICAL MEMORY (PM), is of central importance. In Baddeley's model PM is handled by the phonological loop, and this element has been assumed to play a central role in L1 development (Baddeley 1986). PM is responsible for the processing and temporary retention of both familiar and novel phonological information. From the outset, the role of PM in the acquisition of NEW words has been stressed. Baddeley, Gathercole & Papagno (1998), in a widely-cited review, present compelling evidence that the ability to recall non-words (as measured in the Non-Word Repetition Test, see below) is a stable predictor of vocabulary development in early childhood and assert that 'the function of the phonological loop is not to remember familiar words but to help learn new words' (p. 158). In addition to L1 vocabulary learning, PM has also been related to spoken language development more generally. Children with greater PM capacity produce utterances of greater length and narratives of increased grammatical and semantic complexity than their lower capacity counterparts (Adams & Gathercole 1996).

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

PM has also been shown to play a central role in child and adult L2 development. The role of PM in learning new sound patterns has been shown to be critical to L2 vocabulary learning (Service & Kohonen 1995; Cheung 1996; Baddeley et al. 1998; French & O'Brien 2008). It has also been implicated in the learning of L2 grammar (Speciale, Ellis & Bywater 2004) and L2 oral performance (Ellis & Sinclair 1996; O'Brien et al. 2006). PM has also been identified as a key component of language aptitude (Sawyer & Ranta 2001; Robinson 2005; Kormos & Sáfár 2008; Hummel 2009). This research will be examined in Part 3.

Visual and spatial memory has received less attention from language researchers, reflecting in part the dominant role that phonological information plays in processing spoken and written language. Phonological information is especially important in alphabet-based languages, in which most WMC research has been carried out in the L2 (a notable exception is early work on WMC in Japanese by Osaka & Osaka (1992) and Osaka, Osaka & Groner (1993)). Recent research, though, has highlighted the importance of visual-orthographic skills in processing written Chinese, and increasing work in this area is expected (Tong & McBride-Chang 2010).

The most recent addition to the Baddeley model is the episodic buffer. The construct of 'episodic memory' is not new (see papers in Baddeley, Aggleton & Conway 2002) and its relatively recent inclusion in the model was motivated by the need to explain adequate language performance by individuals with significant phonological loop deficits. These individuals fail, or do very poorly, on tests that measure phonological memory and also have difficulty learning new material. However, language performance is not completely impaired, as they are able to recall narratives and even groups of playing cards that have already passed in games such as contract bridge. It has been posited that the episodic buffer permits this information to be retained and reused (Baddeley et al. 2002). To date there has been no research on the possible role of the episodic buffer in L2 learning and use.

The three short-term storage mechanisms are controlled by the central executive. This control takes several forms. The central executive binds information from a number of sources into a coherent episode by coordinating the working of the various stores (Baddeley 1986). In the performance of complex cognitive tasks it also controls the continuous shifts between immediate task performance and the retrieval processes needed to carry out the task. Crucially, it controls the selective attention needed to maintain focus and inhibit information that might distract from or interfere with successful task execution (Engle & Kane 2004). The control of attention is the primary function of the central executive in all the approaches to WM that have been adapted for use in L2 research (e.g. Baddeley & Logie 1999; Engle & Kane 2004; Cowan 2005). For some researchers, like Engle & Kane (2004), attentional control is synonymous with WM. The relationship between attentional demands in L2 tasks and WM capacity is receiving increasing attention by L2 researchers (Ellis 2001; Mackey 2006b; Rai et al. in press). The Baddeley model of WM is not the only one that is available (see also Ericsson & Kintsch 1995 and Cowan 2005), but it has been the most influential.

As the foregoing indicates, WM is directly implicated in a wide range of language processes, with PM and executive control functions being particularly important elements in language (L1 and L2) learning and use. In addition, WMC has been shown to correlate highly with standard measures of intelligence. This has led some to propose that WM IS general intelligence, as measured in the IQ literature (Colom et al. 2006b). Others, however, while

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

recognizing the correlation between the two, argue that they are separate constructs (Conway, Kane & Engle 2003; Ackerman, Beier & Boyle 2005). Regardless of this debate, the role of WM in language performance, L1 and L2, child and adult, is well established. We will be examining this role in specific L2 domains in Part 3, but first we will describe the methodology used to measure WM capacity and processes in the L2.

2.2. Measuring working memory capacity

Working memory capacity is determined by both storage and processing components, which can be measured separately or in combination. Simple short-term storage capacity is typically measured by the number – or SPAN – of unrelated digits or words that can be recalled. Processing capacity is measured using tasks that make simultaneous demands on storage and processing, sometimes referred to as complex WM (Colom et al. 2006a). The most well known of these is the Reading Span test (Daneman & Carpenter 1980), which is described below. Although short-term storage and complex WM have been treated as separate entities (Harrington & Sawyer 1992; Miyake & Shah 1999), the relationship between the two remains an open empirical question, with evidence both for and against the notion that both elements tap the same underlying system. This issue is taken up in 2.4.

2.2.1. Short-term memory tests

Short-term memory capacity is measured by word and digit span tests that involve the recall of sets of unrelated words or numbers, presented in either written or aural mode (Shah & Miyake 1996). Items are usually presented in sets of ascending size, with the individual recalling increasing numbers of items until a maximum is reached. The use of the word span test assumes prior knowledge of the word on the part of the individual, thus confounding to some extent PM capacity and language knowledge (Gathercole 1995). This is, of course, a concern in testing L2 learners, who may have limited or no familiarity with the L2 words used. The effect of previous language knowledge can be lessened by using digits instead of words (Harrington & Sawyer 1992).

However, a digit span test still requires some knowledge of the language. One way to lessen this problem is to use a non-word repetition task (NWR), in which nonsense words are presented and recalled (Baddeley et al. 1998). The non-words can conform to the phonotactic rules of the language or include sounds that are not in the language (Gathercole 2006). They can also vary in how much they resemble actual words in the language. Testing non-words that include sounds that are not in the first language of the child or adult learner provides a way to assess the individual's ability to encode new phonological sequences.

NWR performance depends on the ability to perceive, store, recall and reproduce phonological sequences, which are all important elements in word learning. The predictive power of performance on NWR in relation to child word learning has been established (Baddeley et al. 1998). Performance on the NWR task has also been linked to word learning

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

in children with specific language impairment (Gathercole 2006). The NWR test has also been shown to predict L2 performance in various domains. These include vocabulary learning in the classroom (Service 1992; Service & Kohonen 1995) and experimental studies (Cheung 1996; Speciale et al. 2004), in differences in oral production (O'Brien et al. 2006; Sunderman & Kroll 2009) and in achievement test performance (Hummel 2009).

2.2.2. Complex working memory tests

As noted, a distinction is often made between simple span tasks that primarily tap storage, and complex tasks that involve both storage and processing. A widely used complex span test is the Reading Span test (Daneman & Carpenter 1980). As with the simple span tasks, variation and controversy has surrounded the Reading Span test (RST) and the constructs it seeks to reveal (Just, Carpenter & Keller 1996; Waters & Caplan 1996a, b, c; MacDonald & Christiansen 2002). The RST assesses an individual's ability to simultaneously read and comprehend a set of sentences and then recall a target word for each, usually the last word in the sentence. Sentences are presented individually in sets ranging from two to six sentences. At the end of each set the participant is asked to recall as many words as possible. Depending on the study, the words can be recalled freely or in the order of presentation, and can be scored in various ways (scoring and other methodology issues are discussed below). In addition to recalling the target words, a secondary comprehension task is also performed. This can involve judging the semantic sense or grammaticality of the sentence. A spoken version of the RST, the Listening Span test, follows the same format but requires the individual to listen rather than read the sentences (Daneman & Carpenter 1980; Mackey et al. 2002; Mackey et al. 2010).

Comprehending the sentences in the Reading and Listening Span tests requires knowledge of the language. A complex span test that lessens the demand on language comprehension is the Operation Span test introduced by Turner & Engle (1989). In this test, simple arithmetic equations are used in place of sentences. Participants solve mathematical operations while trying to remember the final words that are paired with each equation. The dependence on prior language knowledge can be further lessened by using letters, rather than words, as the target recall material (Conway et al. 2005). There are a number of variations of the complex span task that have been used in experimental and clinical research. These include alphabet span, running item span and other tasks involving letters and digits (Waters & Caplan 2003).

An alternative to the complex span tasks is the Backward Digit Span task (Kormos & Sáfár 2008). In this task a sequence of spoken random digits are presented and the participant in the study repeats the numbers in reverse order. The task further minimizes the role of language knowledge in that digits, and not words, are recalled. It also differs from the other span tasks in that the material processed in the secondary task, which is the transformation of the digit order, is the same as the material recalled. This is in contrast to the RST, where the sentences read for comprehension are not recalled (Conway et al. 2005). The test is used widely in neuroscience research (Oberauer et al. 2003). Test performance has been strongly implicated in measures of general intelligence (Engle, Kane & Tuholsky 1999) and is included in the Wechlser IV intelligence test (Gathercole & Alloway 2008). Both the operation span

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

and backward digit span tests avoid the problems arising from including an overt language component.

2.3. Methodological issues in the measurement of working memory

The accurate measurement of WM capacity presents a number of challenges, especially when undertaken outside a controlled laboratory setting. Since the early research (Harrington & Sawyer 1992; Service 1992), L2 studies of WMC have usually been carried out in the classroom or language program setting, where participants are often tested as groups. Memory tests are typically administered along with other tests that measure some aspect of L2 learning or use, with performance across the various measures then being correlated to test a particular research question. A number of methodological factors – how the memory tests are administered and interpreted – can affect the validity and reliability of the findings. The significant factors concern how the memory tests are presented, the nature of the secondary processing task in complex measures like the RST, and the interpretation of the quantitative, usually correlational, results.

For the results to be both valid and reliable, the simple and complex memory tests require that the individual is operating at optimum capacity and that there is no scope for strategic processing that might enhance performance in a way that is not directly related to memory capacity. This requires that the tests are presented in such a way that the individual responds immediately. In the complex tests (RST, Backward Digit Span) it is important that both the processing and storage components are engaged. In the RST or Operation Span test, the secondary processing task must interfere with rehearsal, to prevent the task from becoming simply a measure of short-term memory. This is done by having participants read the sentence aloud and presenting successive items immediately.

The RST can be administered individually or in groups. The latter format has been particularly common in L2 research for practical reasons. However, group presentation makes it more difficult to ensure that all individuals are involved in the processing task. It is also possible that those who can carry out the concurrent processing task more quickly might have more time to (covertly) rehearse the target words. Such covert rehearsal is a confound because it introduces strategies such as creating a story with the set of unrelated words into the task in order to create a relationship and thus better recall them. Test presentation can also be driven by the tester (or computer) or by the test-taker. Friedman & Miyake (2004: 153) note that 'drastically' different results can arise from different administrations of the test.

The nature of the concurrent task is also an issue. Grammaticality judgments and sentence verification tasks involving judgments of truthfulness or logicality have been used. Abu-Rabia (2001) used a Reading Span test that incorporated a sentence completion task. Instead of memorizing the final word, test-takers have to provide a missing word at the end of each sentence, and then recall all the missing words. What is important is that the secondary task draws on processing resources and allows the researcher to establish that it is actually being carried out in the manner intended (e.g. sentence comprehension or arithmetic calculation). A systematic study has yet to be undertaken that compares the various approaches as reliable measures of WMC.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

WMC is reflected in the test scores. Scoring in complex span tests is complicated by the fact that the storage component and the processing component can both contribute to measurement. The relative contribution that processing and storage domains make to RST performance, as well as how the outcomes are interpreted, is subject to debate (Waters & Caplan 2003; Friedman & Miyake 2004). The original version of the RST was carried out manually and scored using an absolute span measure. Sentences were printed on 3 × 5 inch cards and were presented in sets of two, three, four, five and six. Participants were instructed to read each sentence on the card aloud, with normal intonation and speed. The experimenter turned over each card as soon as the participant finished reading the previous one. Once the participants had read each set of sentences, their task was to remember all the words in that set. The only condition was that participants could not mention the word in the last sentence in the set as the first one that they remembered. They were told not to emphasize the last word and that they had to wait until seeing the last card before beginning to recall the list of words. Once the participant failed more than three of the total set of five at a particular level, the test was stopped. If participants got two out of the five sets correct, they were awarded a half point for that level. Studies that have used absolute span scoring in L2 research are Osaka, Osaka & Groner (1993) and Sagarra (2000). However, this results in a very restricted range of score values, usually from two to six, and thus a loss of sensitivity (Oberauer & Süß 2000). This method of administering the task has been abandoned since the early 2000s in most work on RST studies.

Current best practice is for the participant to do the RST task individually on a computer carefully supervised by an experimenter (Conway et al. 2005; Rodríguez 2008) and to use an alternative to absolute span scoring. In this scoring method, credit is given for each correct word recalled, a comprehension/processing check is made, and a reaction time is recorded (Harrington & Sawyer 1992; Miyake & Friedman 1998; Mackey et al. 2002). This can be done either ignoring any errors on the secondary processing task or scoring only those responses from correct items on the comprehension probe. This partial credit scoring has been shown to be superior to the absolute span technique in terms of providing a finer discrimination between individuals and being more reliable, that is, having greater internal consistency (Conway et al. 2005).

However, despite the widespread recognition that the RST as a TASK combines both storage and processing, the recall score (absolute or partial) reflects the output of the storage function. Waters & Caplan (1996b) therefore introduced a modification of the original Daneman & Carpenter (1980) task in which reaction times and accuracy scores, as well as recall scores, were measured for the sentence-reading component. Walter (2006) adopted this technique in the study of cross-linguistic reading transfer. The WMC score in the study consisted of three components, the recall scores (scored by the absolute span method), the percentage of sentences correctly judged as logical, and sentence reading times. The WMC score represented the average of the standardized values of the three measures and was to have strong correlation with text comprehension tasks, with correlations ranging from .49 to .79 across learner levels and tasks. As only the correlations for the overall WMC measures were reported, it was not evident whether the inclusion of accuracy scores and reading times improved the predictive power of the original recall measure, and if so, how. However, the inclusion of specific measures of processing in L2 WMC studies does deserve more attention.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Finally, it is important to note that L2 WM research is principally correlational in nature, and that the standard cautions for interpreting correlational data also apply here. The first of these concerns the obtained strength of association (high, medium or weak). In Kormos & Sáfár (2008), learner performance on the backward digit span test was correlated with scores on the Cambridge First Certificate Examination (FCE). An overall correlation between the two measures of $r = .55$ was reported, while the correlation between Backward Digit Span measure and FCE subscores varied from .19 for writing and .47 for usage. The strongest correlation still accounts for only some of the variance between the two variables (here approximately a quarter). The significance of a reported effect size must therefore be interpreted within the context of the study. The second concern relates to the sensitivity of correlations to sample size. As sample size increases, the likelihood of obtaining a statistically significant correlation also increases. Thus the absence of a significant correlation, especially in comparison to previous studies, may merely reflect different sample sizes. It may also reflect the composition of the sample, as correlations are sensitive to RANGES of scores. If a group of learners are too closely grouped in terms of ability, it is likely that no correlation will be evident.

Finally, the usefulness of a set of scores depends in large measure on the reliability, or internal consistency, of the responses to the items. The reliability of some commonly used tests has been well established in L1 populations (Lezak, Howieson & Loring 2004). Reliability is measured in term of internal consistency, and less commonly, test-retest reliability (Waters & Caplan 2003). Internal consistency is measured either through a measure like Cronbach's alpha, which assesses inter-correlation of items, or split-half reliability, which compares individual performance across two parts of the test. Waters & Caplan (2003: 551) reports split-half reliability coefficients ranging between 0.7 and .09 for Reading and Operation Span measures from a number of published studies. L2 articles rarely report reliability values (Hummel (2009) is an exception). Reporting reliabilities will provide a better picture of performance and, in particular, will help in the interpretation of the reported correlations. The correlation of two measures is limited by the reliability of the respective measures. This is because the correlation between any two measures cannot exceed the product of their reliabilities (Conway et al. 2005: 776). Thus a lack of correlation could reflect the lack of reliability of one or both of the variables, or it could indicate that in fact there is no discernable association between the two.

In this section we have introduced the WM construct as it has been developed in the cognitive psychology literature, and its application in L2 research. We described the key WM tests used in L2 research, including the Reading Span test, and identified important methodological considerations in the use of these tests. In Part 3 we will examine how individual differences in WM capacity have been implicated in various domains of L2 learning and use.

3. Applications in SLA and pedagogy

The constructs of PM and WM span and their associated tests described in Part 2 have been applied to knowledge of SLA and L2 pedagogy in a variety of settings. Because SLA and

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

L2 pedagogy seek to address the question of why some learners struggle to acquire an L2 (usually in instructed contexts), researchers have been interested in finding out whether WM limitations explain differences in success in a variety of domains. The assumption is, of course, that higher WM will lead to more successful learning. As with most broad questions of this nature, breaking down the problem into smaller sub-questions first is more likely to lead to results that make sense. The caveat with all of this research is that although WM may indeed be a factor in explaining SOME variability among learners, other factors such as L1 and motivation may turn out in the end to be much more powerful explanatory variables in L2 learning. Moreover, if it turns out that WM does play a role, AND WM is a fixed TRAIT (perhaps in specific domains) rather than a capacity or attentional variable that can be changed or manipulated, then even knowing that WM is an explanatory variable may not help in pedagogy. Our current state of knowledge is such that we do not know for sure what amount of variation in an individual's learner behavior WM can explain, so researchers have not yet attempted to develop interventions that may help learners increase their WM capacity and, as a result, achieve greater proficiency, or more rapid learning, in instructed contexts.

In our search of the literature we came across many studies that relate to several aspects of WM in a variety of domains. Within the limits of a journal article, we have had to omit some research areas that are potentially relevant. In our choice of sources, we also concentrated mainly on work published in refereed journals rather than book chapters, but we do refer to some dissertations. Given limitations of space, we have not been able to review every study in detail, but we have tried to draw out and illustrate themes from the literature, sometimes by referring to the same study in different contexts.

The main application of the RST has been in studies that seek to explore individual differences in sentence processing, particularly in contexts in which learners are asked to resolve long-distance dependencies in morpho-syntax or resolve complex syntactic structures that are temporarily ambiguous. In a sense, these studies can be called 'reading', but in fact they address precise questions with regard to processing rather than global 'reading skill' as the term is generally understood in pedagogy. These studies are summarized and described in Part 3.1. The RST is the most appropriate for reading and syntactic processing studies (Daneman & Merikle 1996) in contrast to tests of PM, which appear to be more appropriate for investigating lexical development and general proficiency.

In Part 3.2, we review a variety of studies using span tasks that have been used principally to explore vocabulary knowledge and 'grammar' in a variety of tests. Such studies seek to discover whether superior phonological processing ability leads to faster and more accurate LEXICAL development.

Part 3.3 reports on speaking and interaction studies that have explored whether learners with a higher WM capacity produce more either in speaking or modified output. Presumably, if learners speak more, they will interact more and, *ceteris paribus*, learn more, because the interaction models of L2 acquisition hold that acquisition is driven by communication breakdown and negotiation (Mackey et al. 2010). Part 3.4 reports on reading and writing studies that investigate the role of WM, and 3.5 addresses the relationship of WM to measures of global achievement. Part 3.6 summarizes studies on WM and study abroad, and 3.7 deals with WM and aptitude.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

3.1. Working memory and sentence processing

In our search of the literature from 1995 to 2010, we found that most research articles address, to a greater or lesser extent, the role of WM in L2 processing of morpho-syntax. The number of papers in this category far exceeds that in any other category discussed in Parts 3.2–3.7. It is indicative of the interest in this topic among theoretical linguists and psycholinguists, rather than researchers interested in pedagogy.

Studies of sentence processing are important in SLA because it is generally held that for acquisition in either L1 or L2 to occur learners must process both the form and meaning of input (e.g. Fodor 1998; VanPatten 2004). Thus, the more WM you have, presumably the more attentional resources you process and the more you can potentially learn, because you have enough memory to process form TOGETHER with meaning. Studies by VanPatten and colleagues have found that unless required to, learners will process meaning WITHOUT form, leading to the undesirable result that they comprehend L2 input through pragmatics and some basic word order, but do not acquire the detailed morpho-syntactic structures used to convey that meaning. This much seems uncontroversial.

However, the trail of research that leads to the use of WM as a variable in SLA is actually more indirect, so in order to understand the role of WM in SLA research a little history is necessary. Juffs & Harrington (1995, 1996) were among the first published papers that used psycholinguistic techniques of self-paced reading in L2 sentence processing. This technique was used because research at the time was focused on whether L2 learners have access to abstract principles of Universal Grammar. In particular, researchers were interested in whether L2 learners knew that long distance wh-questions such as 'Who do you think __ will win the election?' are possible, whereas ungrammatical questions such as '*What did you meet the politician who promised __?' are not. Although learners were very good at rejecting ungrammatical questions, it turned out that they were less accurate in accepting grammatical sentences, especially ones where the subject has been moved to the front of the sentence, as in the example provided. White & Juffs (1998) and Juffs & Harrington (1995) found that reading sentences with subjects moved from an embedded clause to sentence-initial position resulted in globally slower reading times than matched sentences with an object moved to the front of the clause. Juffs & Harrington (1995) also showed that the increased reading time was on the verb in the EMBEDDED clause, supporting a view that wh-gap resolution was much more difficult in subject extraction than object extraction. The question that then arises is whether the processing is easier if you have more WM – the prediction is that L2 learners with higher WM should be more able to deal with such complex syntax.

The Reading Span task (Daneman & Carpenter 1980) is the most appropriate type of measure of WM in the case of L2 processing research. As described in Part 1, the Daneman & Carpenter (1980) WM measure (RST) is the foundation of a large research literature on the psychology of reading and comprehension for adults (see Conway et al. 2005). As far as we are aware, RST measures have rarely been used to track first language DEVELOPMENT in children (Adams, Bourke & Willis 1999 is, to our knowledge, a rare exception), probably because the tasks are demanding, and because very young children cannot read (Ben-Yehudah & Fiez 2007). Since the introduction of the test in 1980, Just, Carpenter and colleagues (Just &

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Carpenter 1992; Just et al. 1996) have developed the Constrained Capacity Model to explain individual differences in reading comprehension, speed and accuracy in resolving ambiguous sentences (King & Just 1991; MacDonald, Just & Carpenter 1992). A striking example has been reported of the effect of differences in RST on the reading and processing of individual sentences. Based on on-line experiments with eye-tracking and self-paced reading (as well as off-line experiments), a consensus has developed that reading involves incremental processing. Essentially, this claim entails that a native-speaker reader of an alphabetical script such as English, Dutch or French does not 'take in' a large amount (say 7–10 words of text) and then decide the appropriate syntax for that set of words. Rather, each word is processed rapidly, and the reader makes split-second assumptions IMMEDIATELY about a possible syntactic structure for that word and those that follow. This view accounts for readers being misled by ambiguous sentences, and the subsequent 'surprise' when their reading goes off track because the structure they had assumed turns out to be wrong. This 'surprise' is known as the Garden Path (GP) effect (Pritchett 1988). This incremental processing theory emphasizes structural, cue-based and pragmatic principles in its account of the resolution of ambiguity, but also allows a role for frequency effects (see recent papers by Roberts & Gibson (2003) and colleagues, as well as Frazier & Clifton (1996), and MacDonald and colleagues listed in the references).

An interesting facet of WM capacity in this model is that the effects of individual memory differences are not fixed, but task-dependent (Just et al. 1996; Miyake & Friedman 1998). For example, a high-memory-capacity individual will be more accurate in comprehension and resolve an ambiguity at crucial points in reading a sentence such as (1) more quickly than a low capacity individual.

(1) The evidence *examined* by the lawyer convinced the jury.

In (1) the verb 'examined' is TEMPORARILY ambiguous between a main verb and a reduced relative clause structure. Pragmatic information may be used to quickly resolve the parse in favor of a reduced relative clause reading because 'evidence' is inanimate and unlikely to be the agent of any 'examining'. High WM capacity readers are able to resolve this ambiguity more quickly than low WM capacity readers. According to Just and colleagues, this is because high capacity readers are able to combine pragmatic and syntactic information in parsing more efficiently than low span readers. On the other hand, in a sentence such as (2), while high capacity readers are more accurate in comprehension, they take MORE TIME to resolve the parse:

(2) The soldiers *warned* during the midnight raid attacked after midnight.

The account of the difference in processing speed between (1) and (2) for high WM capacity readers is that in (1), high WM individuals are able to make rapid use of pragmatic information, whereas in (2) the ambiguity of 'warned' sets up three PURELY SYNTACTIC possible parses: a main verb reading, an intransitive verb reading, and a reduced relative reading. Just and colleagues argue that in this case high WM individuals are able to maintain all three parses active in parallel, and hence take LONGER to process them. Ultimately, however, they

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

are more accurate with comprehension probes, whereas low WM capacity individuals are faster, but less accurate. Low WM individuals allow the parse to crash, and therefore read more quickly. However, the cost is that they reject these sentences as implausible or fail to understand the relationships among the noun phrases.

The model also relates to differences in scores on standardized tests such as the American Scholastic Aptitude Test (SAT). The SAT is a test conducted in the United States that assesses academic preparedness for university study (e.g. Daneman & Hannon 2001). Their findings suggest that the higher one's RST the better one scores on these tests.

Based on this background, in a series of papers Juffs (2004, 2005 and 2006a) explored whether WM capacity could explain individual L2 learner differences in reading speed in structures that impose very high processing cost. The structures that these papers investigated were Garden Path sentences such as 'After the students ate the pizza TURNED out to be bad', long distance wh-movement such as 'Who does the nurse know __ SAW the doctor in the office?' and sentences that contain temporary ambiguity between a main verb and a reduced relative clause, e.g. 'The bad boys watched every single day WERE playing soccer in the park'. In the self-paced reading experimental technique, words are read one at a time with every press of a computer key. Reading times at key points in the sentences ranged from 400ms to over 1200ms in some cases for non-native speakers. Reading times on verbs, such as those in small capital letters in the examples, can produce very high reading times because readers have to determine who is doing what to whom in the action described by the verb. Because processing pressure is so high here, effects of WM are expected and have been reliably found in L1 self-paced reading studies (Just & Carpenter 1992). However, in most cases, no reliable evidence for the role of WM as measured by a Reading Span task was found in Juffs' studies, but reliable effects of L1 were found in the L2 processing. Japanese speakers, who have verb-final syntax and no wh-movement, experienced the most processing difficulty, followed by the Chinese speakers. It was the Spanish-speaking participants in the study who were able to process such difficult structures with the greatest ease.

It is possible that Juffs' use of the absolute span method and manual method of conducting the span tasks in L1 and L2 led to the failure to find reliable relationships within groups because there was an insufficiently large range of scores. Juffs measured WM in both L1 and L2 (avoiding a confound with L2 proficiency) but still found no reliable effect in either L1 or L2 for any of the groups. Moreover, he found uneven relationships between the word span tasks and the reading span tasks, which is not predicted by some models that treat PM and RST WM as separate constructs.

More recently, Rodríguez (2008) examined the relationship between WM and three structures that impose high processing load – anaphoric reference, long distance relative clause resolution and verb-based ambiguity resolution – to highly proficient Spanish-speaking and Chinese-speaking learners of English. Rodríguez' scoring and testing method for the span tasks followed Conway et al. (2005) very closely, and he also administered a Spanish reading span task to his Spanish participants. However, in none of these studies did he find an effect of WM on reading times at key points in the sentence. However, the first language – Spanish or Chinese – did have a significant effect. No results for the Chinese L1 WM are reported, since he did not have a Chinese language WM test.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Independently of research in Juffs' laboratory, Felser & Roberts (2007) examined the role of WM in explaining differences in a complex cross-modal task (using pictures and reading) in their study of Greek-speaking learners of English and processing of complex relative clauses. They also found no effects for WM.

In contrast, Dussias & Piñar (2010) reinvestigated structures tested in Juffs & Harrington (1995) and Juffs (2005). They used similar linguistic stimuli to Juffs & Harrington (1995) with wh-extraction and Chinese-speaking learners of English L2, but also manipulated plausibility. Dussias & Piñar found that high span learners, but not low span learners, were sensitive to subject/object asymmetries and plausibility. Dussias & Piñar used a version of the Waters & Caplan (1996b) Reading Span test described in Part 2, which means that the test was in English. There is therefore a potential confound of WM test in English and participants' English proficiency, but the different results Dussias & Piñar found for WM clearly shows how the methodology used to administer and score the test can affect results.

This set of studies implies that WM as measured by a Reading Span task MAY not be detectable because the role of the L1 is greater in self-paced reading, especially at lower levels of proficiency.

In contrast, a role for WM in complex syntactic processing has been found in other studies. However, the role proposed varies somewhat by study and is never said to account for a large part of the variation found in individual performance. In most cases, researchers report that the effects of WM are mediated by the first language, as in Juffs' studies, or by pragmatic information. For example, Havik et al. (2009) investigated German-speaking learners of Dutch in a self-paced reading study that involved object and subject relative clauses. Their results indicate that high span non-native speakers are able to use their WM capacity to come close to the behavior of low span NATIVE speakers. However, they found that when only structural information was available to guide parsing decisions, WM span had no effect on the L2 learners' on-line processing, and this differed from the native speakers' on-line processing even though German and Dutch are obviously closed related languages and form relative clauses in a similar way. This finding indicates that WM cannot compensate for processing PURELY structural difficulties presented by an L2 in this self-paced reading methodology.

WM measures have also recently been used to explain individual differences in processing in L2 Spanish among college-age learners in college language programs (e.g. Sagarra 2007a, b). Most recently, Sagarra & Herschensohn (2010) tested the hypothesis that English-speaking learners of L2 Spanish will be unable to process gender agreement features in noun phrases with determiners and agreement because English does not have noun classes, whereas they will be able to process number features because English does have number agreement. The paper also addresses the issue of WM as measured by an adapted span task from Waters & Caplan (1996b). Crucially, this task was in the participants' L1, not Spanish, so the problem of a confound between WM and L2 proficiency is avoided. The basic findings are that the beginning learners are insensitive to gender agreement errors, whereas the intermediate learners show incipient sensitivity to agreement errors in on-line processing. The only effects found for WM were among the higher intermediate learners on ACCURACY measures rather than PROCESSING measures. This lack of effect during processing with the time-course of self-paced reading is consistent with the work of Juffs (2005, 2006a) with ESL learners.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

These results from L2 studies contrast strongly with earlier findings for first language summarized in Just et al. (1996). In the constrained capacity model of reading proposed by Just, Carpenter and their colleagues, millisecond-by-millisecond processing decisions can vary among individuals based on their reading span and the task involved. In L1 processing, high WM readers have an advantage in that they can integrate pragmatic information in order to resolve ambiguity, and high WM will allow multiple parallel syntactic analyses for purely syntactic ambiguity. In L2 processing, the L1 overwhelms the individual differences within language groups based on WM capacity. Hence, the within-language-group individual variation is less important than the between-group-language variation in L2 processing.

3.2. Phonological memory, working memory and vocabulary learning

As noted in Part 2, phonological short-term memory has been shown to play an important role in language learning, especially in the development of vocabulary learning in L1 children. Since the original research was published (Baddeley et al. 1998), the relationship between vocabulary learning and phonological short-term memory, particularly as measured by a non-word repetition test, has often been replicated in L1 and L2 and in cases of specific language impairment (SLI) (Gathercole 2006).

Service and her colleagues established a link between NWR test-performance on the learning of English vocabulary by Finnish schoolchildren (Service 1992; Service & Kohonen 1995). Cheung (1996) reported similar effects for Chinese high school students, but only among the lower proficiency group. Papagno & Vallar (1995) showed a link between PM and vocabulary learning, as did Masoura & Gathercole (2005). In a widely cited study, Speciale, Ellis & Bywater (2004) demonstrated the importance of phonological short-term store capacity. They claim that learners' ability to learn phonological sequences is a predictor of the ability to learn novel L2 vocabulary both in an experimental setting and during a 10-week university Spanish course. Students' initial skill in phonological sequence learning predicted their final levels of Spanish recognition vocabulary and also their ability to repeat Spanish word-like non-words. The authors argued that phonological store capacity and phonological sequence learning ability are initially separable constraints on L2 vocabulary acquisition, and conclude that vocabulary learning is constrained both by limits on phonological short-term store capacity and on the learner's development of the phonological regularities of language. O'Brien et al. (2006) reported similar findings for spoken vocabulary.

Two findings emerge from the foregoing. The first is that, for L2 learners, the role of phonological short-term memory is different, and perhaps greater, for less proficient learners (Cheung 1996; O'Brien et al. 2006; Kormos & Sáfár 2008). The second is that the effect of phonological short-term memory on vocabulary development interacts with long-term knowledge in the developing system (Abu-Rabia 2001; Speciale et al. 2004; Masoura & Gathercole 2005).

However, some studies report no correlation between short-term memory and L2 vocabulary learning. As noted, the more advanced learners in Cheung (1996) and O'Brien et al. (2006) showed no correlation. Other studies have also reported no correlations between word span and vocabulary learning at all (Akamatsu 2008; French & O'Brien 2008).

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

3.3. Working memory and spoken output

Studies that investigate spoken output are beginning to look at the role of WM and spoken production in classroom-based research. Mackey et al. (2002) was among the first studies to look for a role of WM in interaction during language learning tasks. Interaction is seen as a requisite for acquisition as much as processing, since it leads to the sort of communication breakdown that helps learners to notice the gap between what they WANT to say and their ability to EXPRESS what they want to say using the appropriate words and forms. Such noticing through output and negotiation is thought to trigger acquisition through (self) correction and modified output. If WM is linked to noticing, then higher WM should lead to great learner gains. These assumptions were the focus of research in Mackey et al. (2002). They suggested that higher WM capacity learners took longer to reflect on their output and compared it with existing knowledge, whereas learners who had lower WM did not. However, they used a combined score from a variety of WM measures as their estimate of the learners' WM, which may not have been optimal. Mackey at al.'s (2010) research expanded this line of research in a study of 42 English-speaking learners of Spanish. They were given a test of WM, which was a LISTENING span task modeled on a standard Reading Span measure (Waters & Caplan 1996b; Waters & Caplan 2004). In addition, their corpus of reformulations from recordings of interaction on a variety of tasks was collected from transcripts. Different responses were coded according to a typology of responses. Modified output turns were correlated with the results of the WM tasks and a reliable relationship was found between the WM measures and the amount of modified output. WM accounted for 17–18% of the variance in these scores, indicating that other factors were also playing a role in the amount of modified output that the learners produced.

In a study reported by O'Brien et al. (2006), phonological working memory (PM) was found to correlate with the development of oral fluency. Non-word repetition was used as the measure of WM and was used to explore variation in Spanish oral fluency. Fluency was operationalized as pauses and hesitations in the output of the learners, who had studied in two contexts – in the home country and during study abroad. The authors attribute 4.5–9.7% of variance to the differences between the groups to WM.

In contrast to the findings of O'Brien et al. (2006), Mizera (2005) found no relationship between WM capacity and a series of fluency measures in beginning learners of Spanish at the college level. Mizera included an operation span task, a word span task, and a speaking span task in his instruments. In fact, grammatical accuracy was the most significant predictor of fluency as judged by a panel of native Spanish-speaking judges of the narratives created by the students.

3.4. Working memory and 'extensive' reading in L2 and in L2 writing

Part 3.1 reported on a sample of research on syntactic processing, a special kind of L2 reading focusing on bottom-up reading skills. However, bottom-up decoding must, of course, work with top-down processing as well. Top-down processing involves invoking schema and

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

topic knowledge to predict content and assist in bottom-up processing. Tyler (2001) reported that learners and native speakers' use of WM capacity differed when the topic of a reading passage was unavailable. Tyler suggested that non-natives rely more than native speakers on topic for comprehension. More recently, Leeser (2007) carried out a study of L2 reading and noticing of future tense morphology in beginners of L2 Spanish. His findings indicate that WM does play a role in comprehension, but again, knowledge of text topics seemed to be a better predictor of performance than differences in WM. In this case, top-down knowledge was a stronger predictor than bottom-up processing mediated by WM. In a similar vein, Walter (2006), investigating low and high intermediate learners of French L2, found that low intermediate learners did not transfer to L2 reading skills the ability to make links between pronominal forms in the reading tasks. Alptekin & Ercetin (2010) found that learners with higher WM capacity in an adapted Harrington & Sawyer (1992) task performed better at inferential reading.

Overall, then, these studies, along with those reported in Part 3.1, suggest that L2 learners use fewer bottom-up, linguistically fine-grained parsing routines and use fewer WM resources of the kind described in the Just & Carpenter (1992) constrained capacity model. Instead, they rely more on top-down processing to overcome challenges in L2 morpho-syntactic processing (cf. Clahsen & Felser 2006). Top-down knowledge will vary by cultural and educational background, which will co-vary with first language.

Studies that investigate the role of WM with L2 writing tasks are quite limited. This gap in the L2 literature may relate to the lack of precise predictions made by differences in PM and Reading Span tasks about monitored written output. However, Abu-Rabia (2001) investigated the relationship between reading, syntactic, orthographic and WM skills in two orthographically different languages, English and Russian. Participants were first-year BA students of English (aged 25–30) at the University of Haifa who had completed high school in Russia. They were tested individually, in Russian and English, on WM, spelling, oral cloze, visual condition, phonological condition, orthographic skills, word attack and word identification. WM was measured by a version of the Reading Span test in which the last word of the sentence was missing and had to be supplied. The results indicated significant positive correlations within and across languages, except for the orthographic skills and some of the correlations of visual skills. The orthographic skills correlated highly significantly within each language but not across languages. Almost all the other linguistic skills correlated significantly both within and across languages. Multiple regression procedures revealed that phonological and spelling skills in Russian, the L1, were strong predictors of word identification in English, the L2. These results are discussed in the light of the interdependence and the script-dependent hypotheses.

Adams & Guillot (2008) report a study of WM and writing in French–English bilinguals aged 12–15 years. Measures were taken of spelling accuracy, text composition and vocabulary knowledge. The authors found that the students' writing reflected specific memory domains and skill sets. Verbal WM was associated with vocabulary and spelling with PM. Text composition was also predictable by visual WM in English, but few of the predictors were reliable in the French compositions. Some of the WM measures in English predicted composition ability in French; French WM scores were not associated with French compositions. The authors do conclude that because the WM measures

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

are associated in the two languages, WM is independent of language. Kormos & Sáfár (2008) report correlations with writing scores in a global proficiency measure, discussed in Part 3.5.

Studies that focus on the WM resources used in the composition process are rare in the L2 literature although an extensive body of research exists in the literature on dyslexia and L1 writing (Kellogg 2004; Berninger et al. 2008). One recent example from this literature is Olive, Kellogg & Piolat (2008), who report on two experiments that examined whether text composition engages verbal, visual and spatial WM to different degrees. The measures of WM used in the experiments reported in the Olive et al. (2008) experiment are unlike those discussed in detail in Part 2 because they involve composition tasks that occur with concurrent tasks involving the detection of syllables, shapes and patterns on a computer screen (experiment 1) or aurally through headphones (experiment 2). The syllable detection task is designed to tap verbal WM, the shape task (same or different shapes have to be identified as they are presented in sequence) reflects visual memory and the changing position of the same shapes used in the visual task taps spatial memory. Hence, direct comparison of WM effects in these studies with other L2 studies that include writing is not possible. Nevertheless, the findings of the first experiment suggest that when required to perform these tasks visually on a computer at the same time as composing a freehand composition, only the verbal and visual concurrent tasks are negatively affected. The authors suggest that this result indicates a lack of engagement in spatial WM when writing. In a second experiment, the mode of presentation was changed to aural presentation. The syllable task remained the same, but the spatial component was varied by which channel in the headphones the syllables were presented. In this case, only the verbal task had a negative impact on composition. The authors interpret this finding as confirming the lack of a role in spatial WM in the composition process and as supportive of the key role of verbal WM in writing.

3.5. Measure of memory and measures of general proficiency

As indicated in Part 1, PM (the phonological loop) and the RST measure of WMC are constructs that have been proposed to explain individual differences in specific domains. Baddeley and colleagues proposed PM as a language acquisition device in part for the lexicon. The power of PM as measured by a non-word repetition task diminishes as the lexicon of the language is built up. Reading Span has been proposed as a constraint on individual processing and reading that explains variation in comprehension, in particular when processing demands are heavy within a sentence or across discourse.

However, many L2 researchers have sought to use measures of WM to explain GLOBAL differences in ability in the L2 by correlating WM scores with measures of global proficiency. To some extent some of these questions have been addressed by early work reported in Service and Cheung, which has already been discussed, and by more recent work by Service et al. (2003) on the relationship between bilingualism, language skill and WM.

A recent example of the use of WM to explain individual variation in global proficiency is Kormos & Sáfár (2008), which reports a series of correlations that are not entirely consistent with previous research. This study examined individual variation in ESL among

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Hungarian-speaking teenage learners of ESL in a bilingual education intensive English setting. One hundred beginners and 21 students took a non-word repetition test and a subset of the students also took a backward digit span test, measuring their WM capacity in a complex span task. Importantly, the scores on these two tests did not correlate, which the authors suggest is indicative of the separability of the phonological loop and general WM capacity. Moreover, the products of these tests were correlated with scores on a Cambridge First Certificate Examination (FCE), a general proficiency test. Kormos & Sáfár report several interesting findings. First, among the true beginners at the inception of the study ($N = 100$) the non-word span test did not correlate with any subcomponent (reading, writing, listening, speaking or use of English) or overall score on the FCE test. The same result was found for the 21 learners who were 'pre-intermediate' when the study began, with the exception that non-word score was weakly correlated with writing performance and 'use of English'. This result needed explanation, because, as we saw in Part 3.2, PM effects are often most evident in the early stages of learning and diminish as the lexicon increases, since PM is generally judged to facilitate early lexical acquisition (Baddeley et al. 1998: 158). Indeed, Cheung's (1996) failure to find PM effects in a population tested in Hong Kong is attributed to the fading predictive ability of PM as children get older. The authors suggest that the weak correlation found for the pre-intermediate learners on these two tasks was due to their ability to 'retrieve words quickly and efficiently' (p. 268) in a holistically graded writing task that combined knowledge of syntax and vocabulary. As with all correlational studies that run multiple correlations, the potential for finding a significant correlation by chance (a Type I error) increases with the number of tests run, so the authors are correct to say that their interpretation of these significant correlations is speculative. In addition, they suggest that the non-word repetition task is more a measure of implicit learning, whereas the participants in this study learned English through explicit instruction and memorization of rules.

The backwards digit span did correlate for all sections of the test (except writing, the exact opposite of the non-word repetition task) and for overall proficiency ($r = 0.55$). Kormos & Sáfár interpret this result as supporting a role for complex WM in text-based learning and learning that requires ATTENTION to instructions and NOTICING (p. 269). In general, these results indicate at best a weak contribution of PM in classroom-based foreign language instruction and a greater role for a WM test related to attention-explicit knowledge.

Another body of work on general proficiency, this time in elementary schools, has been developed by French and his colleagues in Québec, Canada, asking how far individual differences in elementary school language development might be explained by WM. The research has focused on 8–11 year old children in Québec. French-speaking children in Québec are exposed to English in primary school, but by law they may not receive CONTENT instruction in any language other than French. To promote English language learning in the elementary schools, therefore, many schools have established five-month 'intensive' English programs in which the children focus ONLY on English language study. French (2006) and French & O'Brien (2008) report that WM tasks explain a large amount of variation in the students' learning. For example, French & O'Brien (2008) describe a study in which they gave non-word repetition tasks, using words from Arabic as a 'non-word' repetition task and one in the L2 (English), to over 100 11 year-olds learning English in a five-month intensive program. They used these WM scores to correlate with various measures of

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

morpho-syntactic development on a multiple-choice discrete point test, and receptive and productive vocabulary development. Using hierarchical regression analysis, the authors found that the WM task in the L2 (English) accounted for at least 28% of the variance in the gains in knowledge of morpho-syntax. Interestingly, vocabulary gains were predicted only by pre-test scores on vocabulary and not the WM test itself; moreover, grammatical development was NOT mediated by lexical development. From these results, the authors conclude that for this group WM did not predict gains in vocabulary, because lexical acquisition had already been established. The WM did, however, predict grammar learning, and they assert that these grammar structures were in some sense like 'new vocabulary'. This interpretation is consistent with emergentist or construction grammar views of language learning, but this interpretation needs to be confirmed through further research. Moreover, the grammar test used discrete point multiple choice, so it is hard to gauge the precise effect of WM on the true grammatical development of these children.

3.6. Working memory and study abroad

Our literature search resulted in three papers that included WM as a variable in a study abroad context. These include the O'Brien et al. (2006) paper mentioned in Part 3.5, which factored out study abroad and found effects for PM on fluency. However, using a reading span task based on Waters & Caplan's (1996b) work, Tokowicz, Michael & Kroll (2004) and Sunderman & Kroll (2009) both examine the role of WM in the study abroad context. Tokowicz et al. (2004) found that learners who studied abroad were more likely to use approximate translations; in addition, only those with higher WM were really able to do so well because this area of strategic competence requires multiple items to be maintained in memory simultaneously. Consistent with this result, Sunderman & Kroll (2009) found that in tests of lexical production, learners with low WM scores appeared less accurate on a production task even though they had had a study abroad experience. The authors attribute this effect to the higher WM individuals' ability to simultaneously suppress the L1 influence while at the same time paying attention to the L2 form–meaning mappings. In other words, higher WM provides a learner with the possibility of marshaling attentional resources more effectively. This should in turn lead to more noticing, which is deemed crucial for certain aspects of SLA (Schmidt 1990). The implication is that learners should be helped to suppress the L1 while learning the L2.

3.7. Working memory and aptitude

The robust role that WMC plays in the L2 raises the issue of how far WMC should be considered a TRAIT, that is, a stable individual characteristic that ultimately has a neuro-cognitive basis, as opposed to a dynamic processing STATE whose capacity and efficiency is affected by factors such as domain knowledge and learner goals. Central to addressing this question is understanding which factors affect capacity limitations, and how stable that individual capacity is over time.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

The trait-based perspective on WMC is best represented in research that has examined the relationship between WMC and language aptitude. The language aptitude construct and its role in L2 learning has been discussed widely (Miyake & Friedman 1998; Sawyer & Ranta 2001; Dörnyei 2005; Robinson 2005; Richard et al. 2006; Kormos & Sáfár 2008). Traditional models of language aptitude identified phonetic coding ability involving the use of spelling cues, grammatical sensitivity involving identifying the role of words in sentences, rote learning ability in paired associate learning and inductive learning ability, which reflects overall aptitude (Carroll 1981). Evidence for the stability of aptitude over time is relatively scarce, and while evidence does exist that aptitude is affected by experience, there is also a recognition that language aptitude, at least in the way it is traditionally measured, is a relatively stable construct over time (Skehan 2002: 79).

The relationship between measures of WMC and traditional language aptitude was examined in Hummel (2009), who found no significant correlation between phonological short-term memory as measured by a non-word repetition task and the Modern Language Aptitude Test (MLAT) subcomponents, although both predicted learning overall. When the group was split into high and low groups based on WMC it was found that PM only correlated with the novice group. This is consistent with the L1 research and with Cheung (1996). In contrast, Robinson (2002) found a moderate correlation between the MLAT subcomponents on the complex RST. Kormos & Sáfár (2008) examined the relationship between traditional language aptitude tests, PM (NWR), complex WM (Backward digit span) and attained proficiency in the four language skills. They found no correlation between aptitude for the simple WM task but did find one for the Backward digit span similar to that obtained by Robinson (2002) using the RST.

Although it is clear that individual differences in WMC are generally stable, and provide a means for comparison, it is also the case that WMC can change as a result of experience. Evidence for a correlation between WM capacity and practice (Turley-Ames & Whitfield 2003; Lee & Chen 2005; van den Noort et al. 2006) suggests that WM capacity is a DYNAMIC construct that reflects the learner's knowledge organization and experience, rather than a relatively fixed individual trait (see also Grigorenko, Sternberg & Ehrman 2000; MacDonald & Christiansen 2002; Service et al. 2004). It has been suggested that WMC be viewed as a mixture of trait and state variable, like anxiety, in which stable patterns of performance across tasks are evident, but that performance can be affected at any particular time by a specific task condition (Engle 2007).

It is of some concern that so many different tests are used and that multiple variations of simple span, non-word repetition and Reading Span tasks are used by different researchers for different language domains, since it makes comparison among all these studies rather difficult. Factors such as variations in the items on the tests and whether the participants are supervised or not during the testing make any future meta-analysis highly problematic. In addition, post hoc rationalization of multiple correlations of various tests of WM, regardless of the theoretical domain of knowledge (e.g. vocabulary, lexis or processing), with assorted measures of language ability is not likely to move the field forward, and rather suggests a confirmatory bias in some of the research which we have surveyed. For example, it is not clear why a Reading Span task based on Waters & Caplan (1996b) would NECESSARILY be expected to correlate with lexical development (e.g. Kroll et al. 2002). Although it appears

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

that the Waters & Caplan (1996b) task has provided explanation of variability in this domain and others, the precise nature of this relationship remains unclear. Some researchers suggest that such tasks measure attentional capacity rather than ability to encode novel phonological sequences, as in the non-word span tasks, or capacity to process ambiguity and reanalyze complex syntax in the Reading Span tasks. If what we have been calling WM is in fact a measure of attention, then the field needs to address this terminological issue directly. Progress may be made on the role of WM itself when the measures of WM are more closely matched and justified in terms of the linguistic domain of acquisition that they target.

4. Conclusion

4.1. Overall findings

In some experimental contexts, such as self-paced reading with high-load ambiguity in syntactic processing experiments, the effects of individual differences on memory spans have not been robustly detectable. When effects are detectable, they often appear to be mediated by other factors such as prior knowledge or pragmatic inferences. Listening Span tests do seem to predict output in interaction-based studies, but there are very few such studies, and a great deal of research is necessary to confirm these findings. In general, it seems that the Reading Span test, or Listening Span test – based on Waters & Caplan (1996b) in English – has been the most successful in explaining some variation in studies published so far (e.g. Sunderman & Kroll 2009; Dussias & Piñar 2010; Sagarra & Herschensohn 2010), but it is unclear why at this point this particular test correlates with such a variety of domains (lexical, gender agreement and syntactic processing) in L2 learning. Assuming that the participants in these studies have not had the opportunity for covert rehearsal, this version of the test seems to be the most promising. Further research with this test, to obtain reliability scores and ensure consistent administration of the test, is a promising way forward. Studies that seek to explain variation in grammatical and processing ability would do better to use this kind of test and test of speaking ability rather than PM tests.

Phonological memory tests appear to be very robust predictors of early language development, but the effects of PM on subsequent lexical development seem to be limited and do not predict gains once a basic vocabulary has been established.

Many researchers assume that WM for later learning is in fact a form of attentional control and that this ability to control attentional resources and suppress competing information sources, rather than memory per se, is the key to differences in language learning success. Sunderman & Kroll (2009) suggest that this is what the measure of WM that they use is picking up on.

4.2. Implications for instruction

The implications for classroom instruction are often not clear, and recommendations from the L1 learning literature and the literature on dyslexia and SLI (e.g. Gathercole, Lamont

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

& Alloway 2006; Gathercole & Alloway 2008) are not easily transferable to L2 learning. However, Hummel & French (2010) suggest that communicative language classrooms, with their emphasis on oral-aural learning, might put low WM learners at a disadvantage. They review some literature on individuals with Specific Language Impairment and individuals with Attentional Deficit Hyperactivity Disorders that suggests that attentional resources can be manipulated. They suggest that additional multi-modal support in the form of multimedia learning tools such as online dictionaries and electronic glosses could be provided. They also mention some experimental studies in which online synchronous chat enables low WM learners to produce the same quantity and complexity of output as higher WM learners. They attribute this result to the reduced cognitive burden and increased planning time that the computer assisted language-learning tool provides (Payne & Ross 2005). In addition, increased amounts of visual aid in developing communicative activities (including the learners' own writing) are suggested as tools that can reduce cognitive load and compensate for lower WM capacity. For text-based processing to increase reading ability, Hummel & French suggest that limited amounts of reading aloud could promote some rehearsal and promote the phonological loop component of WM. However, this suggestion has not been supported by empirical research.

If WM is a trait that cannot be changed, it is hard to see how it can be altered to improve language learning. However, if WM is linked to attentional capacity, or control of the Central Executive function, which can suppress competing pressure from the L1, as Sunderman & Kroll (2009) suggest, then learners could be assisted in focusing on L2 forms and attempting to suppress L1 influence. Given that many scholars, whatever their position on the theory of linguistic knowledge, assume that L1 transfer is inevitable, learning to control attentional resources and suppress that influence could be a key to more successful L2 learning. How such suppression may be achieved is a question for future research, although Hummel & French (2010) do refer to studies of 'input flood' (Trahey & White 1993) as potentially relevant in this context.

4.3. Future directions

Several themes emerge from our review of the role of WM in L2 learning and use. It is clear from the literature that 'WM' is not a unitary construct. Rather, it is a set of processes that underpin the learning and use of a second or additional languages. The dual functions of WM (storage and processing) vary in importance according to the L2 domain, and it is clear that the various WM subsystems (PM and/or RST) may change in their importance over time and domain. How these subsystems interact, and the role they play in constraining L2 performance, remains to be answered. It is important to emphasize that the effect of WMC as a constraint on L2 performance will differ by L2 domain, and evidence for its presence or absence in an area (e.g. sentence processing) may tell us little of the role it plays in other, arguably more important, domains (discourse comprehension).

Moreover, a distinction has been made between how WM affects 'cold' and 'hot' cognition (Unsworth, Hertz & Engle 2005). 'Cold' processing refers to more subconscious, 'rational' elements of cognition, whereas 'hot' processing involves affect. The focus in L2 research has

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

been primarily on various aspects of cold processing, which focus on information-processing characterizations of L2. These areas include vocabulary learning (Cheung 1996), sentence processing (Juffs 2005; Dussias & Piñar 2010), grammar learning (Speciale et al. 2004), orthographic development (Abu-Rabia 2001) and discourse comprehension (Walter 2006). However, hot cognition has also received attention. For example, affect/extroversion (Dewaele et al. 2008) and socio-cultural orientation to a task (Al-Namlah, Fernyhough & Meins 2006) have been the focus of recent innovative research.

A second theme that emerges is that the effect of WM on learning differences in individuals changes over the course of development. For example, O'Brien et al. (2006) have shown that PM plays a significant role in early vocabulary and, later, perhaps also on grammar (French & O'Brien 2008). A more nuanced understanding of these changes is needed.

A third theme is that attention is important. WM research may be a way in to a better understanding of the role of attention in SLA. A decade ago Ellis (2001) noted that the role of attention is little understood. One can ask how much progress has been made. WM may be a way to study it, but we need to be careful about the term 'attention' itself. It is used in a very broad sense, often interchangeably with 'noticing' (Schmidt 1990), to refer to behaviors that run a time course that can be measured in seconds, or even tenths of seconds. In the L1 literature, researchers such as Engle and colleagues, who ascribe a central role to attention, discuss attention in terms of processes that involve maintenance of immediate attention and resistance to distraction and that are sensitive at the millisecond level.

A fourth theme is that methodological development is an important characteristic of L2 WMC research. This innovation is reflected in the introduction of new tests, such as the speaking span test (Fortkamp 1999; Mizera 2005; Weissheimer & Mota 2009), the use of different and more engaging comprehension tasks, such as the sentence completion task used by Abu-Rabia (2001), and the response time measures to assess processing more directly (Walter 2006). As L2 researchers employ increasingly sophisticated measures we anticipate a much more nuanced understanding of the role of WMC in L2 learning.

The contribution of phonological memory will continue to be an important area of research. One particularly fruitful avenue is the incorporation of other aspects of phonological knowledge such as phonological awareness and sequence learning using standardized tools (e.g. Wagner, Torgeson & Rashott 1999), to make L2 research as replicable as in the clinical literature (Lezak et al. 2004; Jefferies et al. 2004; Ben-Yehudah & Fiez 2007). This research may be particularly important for beginning learners and also for understanding the literacy challenges of at-risk groups of low literate learners (Juffs 2006b; van der Craats, Kurvers & Young-Scholten 2006). Moreover, while research on phonological memory will continue to be important, we can expect more work on how visual information is processed as part of language comprehension and written output (gesture and gaze research as well as spatial WM, cf. Olive et al. 2008). In this context, a very recent line of research that has compared English–Chinese bilinguals processing English and Chinese orthography is promising (Tong & McBride-Chang 2010). In addition, while work in cognitive neuroscience has begun to explore the role of WM and L2 processing with event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) (Tolentino 2010; Kim et al. 2011), more work is needed to clarify behavioral data before real advances in cognitive neuropsychology can be made.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Much remains to be understood where the various constructs, operationalizations and tests used to measure WM and their relationship to L2 development are concerned. As the field matures, the potential exists for the relation between WM and L2 development to contribute insights to WM studies in general.

Acknowledgements

This research was supported in part by a grant from the Pittsburgh Science of Learning Center to Alan Juffs. The Pittsburgh Science of Learning Center is funded by the National Science Foundation award number SBE-0836012, and was previously funded by NSF award number SBE-0354420.

We are also grateful to our research assistants Katherine I. Martin at Pittsburgh and Lara Weinglass at the University of Queensland for their help in the search for references in journals.

References

Abu-Rabia, S. (2001). Testing the interdependence hypothesis among native adult bilingual Russian–English students. *Journal of Psycholinguistic Research* 30.4, 437–455.

Ackerman, P. L., M. E. Beier & M. Boyle (2005). Working memory and intelligence. *Psychological Bulletin* 131.1, 30–60.

Adams, A.-M. & S. E. Gathercole (1996). Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology* 49.A, 216–233.

Adams, A.-M. & K. Guillot (2008). Working memory and writing in bilingual students. *International Journal of Applied Linguistics* 156, 13–28.

Adams, A.-M., L. Bourke & C. S. Willis (1999). Working memory and spoken language comprehension in young children. *International Journal of Psychology* 34.5/6, 363–373.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics* 9, 175–193.

Al-Namlah, A. S., C. Fernyhough & E. Meins (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology* 42, 117–131.

Alptekin, C. & G. Ercetin (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading* 33, 206–219.

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, 417–423.

Baddeley, A. D. & G. Hitch (1974). Working memory. In G. H. Bower (ed.), *Recent advances in learning and motivation*. New York: Academic Press, 47–89.

Baddeley, A. D. & R. H. Logie (1999). Working memory: The multicomponent model. In A. Miyake & P. Shah (eds.), *Models of working memory*. Cambridge: Cambridge University Press, 28–61.

Baddeley, A. D., S. E. Gathercole & C. Papagno (1998). The phonological loop as a language-learning device. *The Psychological Review* 105, 158–173.

Baddeley A. D., J. P. Aggleton & M. A. Conway (eds.) (2002). *Episodic memory: New directions in research*. New York: Oxford University Press.

Ben-Yehudah, G. & J. A. Fiez (2007). Development of verbal working memory. In D. Coch, K. Fischer & G. Dawson (eds.), *Human behavior and the developing brain* (2nd edn). New York: Guilford Press, 301–328.

Berninger, V. W., W. Raskind, T. Richards, R. Abbott & P. Stock (2008). A multidisciplinary approach to understanding development dyslexia within working memory architecture: Genotypes, phenotypes, brain and instruction. *Developmental Neuropsychology* 33.6. 707–744.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University 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*. Rowley, MA: Newberry House, 119–154.
Cheung, H. (1996). Non-word span as a unique predictor of second language vocabulary learning. *Developmental Psychology* 32, 867–873.
Clahsen, H. & C. Felser (2006). Grammatical processing in language learners. *Applied Psycholinguistics* 27, 3–42.
Colom, R., I. Rebello, F. Abad & P. C. Shih (2006a). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition* 34, 158–171.
Colom, R., P. C. Shih, C. Flores-Mendoza & M. A. Quiroga (2006b). The real relationship between short-term memory and working memory. *Memory* 14, 804–813.
Conway, R. A. (2007). *Variation in working memory*. Oxford: Oxford University Press.
Conway, R. A., M. J. Kane & R. W. Engle (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science* 7, 547–552.
Conway, R. A., M. J. Kane, M. Bunting, D. Hambrick, O. Wilhelm & R. Engle (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12.5, 769–786.
Cowan, N. (2005). *Working memory capacity*. New York: Taylor & Francis.
Daneman, M. & P. A. Carpenter (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19.4, 450–466.
Daneman, M. & B. Hannon (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology General* 130, 208–223.
Daneman, M. & P. M. Merikle (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review* 3, 422–433.
Dewaele, J.-M., K. V. Petrides & A. Furnham (2008). Effect of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning* 58.4, 911–960.
Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
Dörnyei, Z. & P. Skehan (2003). Individual differences in second language learning. In C. Doughty & M. Long (eds.) *The handbook of second language acquisition*. Malden, MA: Blackwell, 539–587.
Dussias, P. E. & P. Piñar (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese–English second language speakers. *Second Language Research* 26, 443–472.
Ellis, N. C. (2001). Memory for language. In P. Robinson (ed.), *Cognition and second language instruction*. New York: Cambridge University Press, 33–68.
Ellis, N. C. & S. G. Sinclair (1996). Working memory in the acquisition of vocabulary and syntax. *Quarterly Journal of Experimental Psychology* 49, 234–250.
Engle, R. W. (2007). Working memory: The mind is richer than the models. In H. L. Roediger, Y. Dudai & S. M. Fitzpatrick (eds.), *Science of memory: Concepts*. Oxford: Oxford University Press, 159–164.
Engle, R. W. & M. J. Kane (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (ed.), *The psychology of learning and motivation*. New York: Elsevier, 145–199.
Engle, R. W., M. J. Kane & S. W. Tuholsky (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press, 102–134.
Ericsson, K. A. & W. Kintsch (1995). Long-term working memory. *The Psychological Review* 102.2, 211–245.
Felser, C. & L. Roberts (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23, 9–36.
Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry* 29, 1–36.
Fortkamp, M. B. M. (1999). Working memory capacity and elements of L2 speech production. *Communication and Cognition* 32, 259–295.
Frazier, L. & C. Clifton (1996). *Construal*. Cambridge, MA: MIT Press.
French, L. M. (2006). *Phonological working memory and L2 acquisition: A developmental study of children learning French in Québec*. Lewiston, NY: Edwin Mellen Press.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

French, L. M. & I. O'Brien (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics* 29, 463–487.
Friedman, N. P. & A. Miyake (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language* 51.1, 136–158.
Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition* 23.1, 83–94.
Gathercole, S. E. (2006). Nonword repetition and word learning: the nature of the relationship. *Applied Psycholinguistics* 27, 513–543.
Gathercole, S. E. & T. P. Alloway (2008). *Working memory and learning: A practical guide for teachers*. Thousand Oaks, CA: Sage.
Gathercole, S. E., E. Lamont & T. P. Alloway (2006). Working memory in the classroom. In S. J. Pickering (ed.), *Working memory and education*. Burlington, MA: Academic Press, 219–240.
Grigorenko, E. L., R. J. Sternberg & M. E. Ehrman (2000). A theory-based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *The Modern Language Journal* 84.3, 390–405.
Harrington, M. W. & M. Sawyer (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition* 14.1, 25–38.
Havik, E., L. Roberts, R. Van Hout, R. Shreuder & M. Haverkort (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning* 59.1, 73–112.
Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in non-novice adult learners. *Applied Psycholinguistics* 30, 225–249.
Hummel, K. M. & L. French (2010). Phonological memory and implications for the second language classroom. *Canadian Modern Language Review* 66.3, 371–391.
Jefferies, E., K. Patterson, R. W. Jones, D. Bateman & M. A. Lambon-Ralph (2004). A category-specific advantage for numbers in verbal short-term memory: Evidence from semantic dementia. *Neuropsychologica* 42, 639–660.
Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society* 102.2, 199–225.
Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21.2, 121–151.
Juffs, A. (2006a). Processing reduced relative vs. main verb ambiguity in English as a Second Language: A replication study with working memory. In R. Slabakova, S. Montrul & P. Prevost (eds.), *Inquiries in linguistic development in honor of Lydia White*. Amsterdam: John Benjamins, 213–232.
Juffs, A. (2006b). Working memory, second language acquisition and low-educated second language and literacy learners. *Low-educated adult second language and literacy acquisition. Proceedings of the Inaugural Symposium*. Netherlands Graduate School of Linguistics, Occasional Series, 89–104.
Juffs, A. & M. W. Harrington (1995). Parsing effects in L2 sentence processing: Subject and object asymmetries in *wh*-extraction. *Studies in Second Language Acquisition* 17.4, 483–516.
Juffs, A. & M. W. Harrington (1996). Garden path sentences and error data in second language sentence processing research. *Language Learning* 46, 286–324.
Just, M. A. & P. Carpenter (1992). A capacity theory of comprehension: Individual differences in working memory. *The Psychological Review* 99, 122–149.
Just, M. A., P. Carpenter & T. A. Keller (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *The Psychological Review* 103, 773–780.
Kellogg, T. L. (2004). Working memory components in written sentence generation. *American Journal of Psychology* 117.3, 341–361.
Kim, K. K., E. Byun, S. K. Lee, W. D. Gaillard, B. Xu & M. H. Theodore (2011). Verbal working memory of Korean-English bilinguals: An fMRI study. *Journal of Neurolinguistics* 24.1, 1–3.
King, J. & M. A. Just (1991). Individual differences in syntactic processing: The rate of working memory. *Journal of Memory and Language* 30.5, 580–602.
Kormos, J. & A. Sáfár (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11, 261–271.
Kroll, J., E. Michael, N. Tokowicz & R. Dufour (2002). The development of lexical fluency in a second language. *Second Language Research* 18, 137–171.
Lee, Y-S. & H-C. Chen (2005). The effect of mnemonic training upon the working memory capacity. *Bulletin of Educational Psychology* 37, 41–59.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning* 57, 229–270.
Lezak, M. D., D. B. Howieson & D. W. Loring (2004). *Neuropsychological assessment* (4th edn). New York: Oxford University Press.
MacDonald, M. C. & M. H. Christiansen (2002). Reassessing working memory: Comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review* 109, 35–54.
MacDonald, M. C., M. A. Just & P. A. Carpenter (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology* 24.1, 55–98.
Mackey, A. (2006a). Feedback, noticing and instructed second language learning. *Applied Linguistics* 27, 405–430.
Mackey, A. (2006b). From introspections, brain scans, and memory tests to the role of social context: Advancing research on interaction and learning (Epilogue). *Studies in Second Language Acquisition* 28.2, 369–379.
Mackey, A., J. Philp, T. Egi, A. Fujii & T. Tatsumi (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (ed.), *Individual differences and instructed language learning*. Philadelphia, PA: John Benjamins, 181–209.
Mackey, A., R. Adams, C. Stafford & P. Winke (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning* 60.3, 501–533.
Masoura, E. & S. E. Gathercole (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory* 13.3/4, 422–429.
Miyake, A. & N. P. Friedman (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 retention and training*. Mahwah, NJ: Lawrence Erlbaum, 339–364.
Miyake, A. & P. Shah (1999). Toward unified theories of working memory. In A. Miyake & P. Shah (eds.), *Models of memory: Mechanisms of active maintenance and recall*. Cambridge: Cambridge University Press. 442–481.
Mizera, G. J. (2005). Working memory and L2 oral fluency. Unpublished Ph.D. dissertation. University of Pittsburgh, Pittsburgh.
Oberauer, K. & H. M. Süß (2000). Working memory and inference: A comment on Jenkins, Myerson, Hale & Fry (1999). *Psychonomic Bulletin & Review* 7, 727–733.
Oberauer, K., H. M. Süß, O. Wilhelm & W. W. Wittman (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence* 31, 167–193.
O'Brien, I., N. Segalowitz, J. Collentine & B. Freed (2006). Phonological memory and lexical narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics* 27, 377–402.
Olive, T., R. T. Kellogg & A. Piolat (2008). Verbal, visual, and spatial working memory demands during text composition. *Applied Psycholinguistics* 29, 669–687.
Osaka, M. & N. Osaka (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society* 30, 287–289.
Osaka, M., N. Osaka & R. Groner (1993). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society* 31, 117–118.
Papagno, C. & G. Vallar (1995). Verbal short term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, 48A, 98–107.
Payne, J. S. & B. M. Ross (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning & Technology* 9, 35–54.
Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language* 64, 539–576.
Rai, M. K., L. C., Loschky, R. J., Harris, N. R. Peck & L. G. Cook (in press). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning*. Article first published online: 18 August 2010, DOI: 10.1111/j.1467-9922.2010.00592.x.
Richard, L. S., P. Jon, G. Leonore, H. Nancy & J. James (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia* 56, 129.
Roberts, R. & E. Gibson (2003). Individual differences in sentence memory. *Journal of Psycholinguistic Research* 31, 573–598.
Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfeld & Hernstadt (1991).

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

In P. Robinson (ed.), *Individual differences and instructed language learning*. Amsterdam: John Benjamins, 211–266.
Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics* 25, 45–73.
Rodríguez, G. A. (2008). Second language sentence processing: Is it fundamentally different? Unpublished Ph.D. Dissertation. University of Pittsburgh, Pittsburgh.
Sagarra, N. (2000). The role of working memory on second language acquisition: A longitudinal study. Unpublished Ph.D. Dissertation. University of Illinois, Urbana-Champaign.
Sagarra, N. (2007a). Working memory and L2 processing of redundant grammatical forms. In Z. Han & E. S. Park (eds.), *Understanding second language process*. Bristol, UK: Multilingual Matters, 133–147.
Sagarra, N. (2007b). Online processing of gender agreement in low proficient English–Spanish late bilinguals. In M. J. Cabrera, J. Camacho, V. Deprez, N, Flores & L. Sanchez (eds.), *Current issues in linguistic theory series*. Amsterdam: John Benjamins, 240–253.
Sagarra, N. & J. Herschensohn (2010). The role of proficiency and working memory in gender and number agreement marking in processing in L1 and L2 Spanish. *Lingua* 120, 2022–2039.
Sawyer, M. & L. Ranta (2001). Aptitude, individual differences, and instructional design. In P. Robinson (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 319–353.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11.2, 129–158.
Service, E. (1992). Phonology, working memory and foreign language learning. *Quarterly Journal of Experimental Psychology* 45.A, 21–50.
Service, E. & V. Kohonen (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16, 155–172.
Service, E., M. Simola, O. Metsanheimo & S. Maury (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology* 14, 383–408.
Shah, P. & A. Miyake (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology General* 125.1, 4–27.
Skehan, P. (2002). Theorising and updating. In P. Robinson (ed.), *Individual differences in instructed language learning*. Amsterdam: John Benjamins, 69–93.
Speciale, G., N. C. Ellis & T. Bywater (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25.2, 293–321.
Sunderman, G. & J. F. Kroll (2009). When study abroad fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics* 30, 79–99.
Tokowicz, N., E. B. Michael & J. F. Kroll (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition* 7, 255–272.
Tolentino, L. C. (2010). Second language Swedish morphosyntactic instruction and cross-language similarity: An ERP investigation. Unpublished Ph.D. dissertation, University of Pittsburgh, Pittsburgh.
Tong, X. & C. McBride-Chang (2010). Chinese-English biscriptal reading: Cognitive component skills across orthographies. *Reading and Writing* 23, 293–310.
Trahey, M. & L. White (1993). Positive evidence and pre-emption in the second language classroom. *Studies in Second Language Acquisition* 15, 181–204.
Turley-Ames, K. J. & M. M. Whitfield (2003). Strategy training and working memory past performance. *Journal of Memory and Language* 49.4, 446–468.
Turner, M. L. & R. W. Engle (1989). Is working memory capacity task dependent? *Journal of Memory and Language* 28.2, 157–154.
Tyler, M. D. (2001). Topic knowledge in nonnative and native comprehension. *Language Learning* 51.2, 257–280.
Unsworth, N., R. P. Hertz & R. W. Engle (2005). Working memory capacity in hot and cold cognition. In R. W. Engle, G. von Hecker Sedek & D. N. McIntosh (eds.), *Cognitive limitations in aging and psychopathology*. New York: Cambridge University Press, 19–43.
Van de Craats, I., J. Kurvers & M. Young-Scholten (2006). Research on low-educated second language and literacy acquisition. *Low-educated adult second language and literacy acquisition. Proceedings of the Inaugural Symposium*. Netherlands Graduate School of Linguistics, Occasional Series, 7–23.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press

Van Den Noort, M. W. M. L., P. Bosch & K. Hugdahl (2006). Foreign language proficiency and working memory capacity. *European Psychologist* 11, 289–296.
VanPatten B. (ed.) (2004). *Processing instruction: Theory, research, commentary*. Mahwah, NJ: Lawrence Erlbaum.
Wagner, R., J. K. Torgesen & C. Rashotte (1999). *CTOPP comprehensive test of phonological processing*. East Moline, IL: LinguiSystems.
Walter, C. (2006). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics* 25.3, 315–339.
Waters, G. S. & D. Caplan (1996a). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition* 24, 342–355.
Waters, G. S. & D. Caplan (1996b). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology-Human Experimental Psychology* 49.A, 51–79.
Waters, G. S. & D. Caplan (1996c). The capacity theory of sentence comprehension: Critique of Just & Carpenter (1992). *Psychological Review* 103.4, 761–772.
Waters, G. S. & D. Caplan (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments & Computers* 35, 550–564.
Waters, G. S. & D. Caplan (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *Quarterly Journal of Experimental Psychology (Section A) – Human Experimental Psychology* 57, 129–163.
Weissheimer, J. & M. B. Mota (2009). Individual differences in working memory capacity and the development of L2 speech production. *Issues in Applied Linguistics* 17, 93–112.
White, L. & A. Juffs (1998). Constraints on Wh-movement in two different contexts of non-native language acquisition: Competence and processing. In S. Flynn, G. Martohardjono & W. O'Neill (eds.), *The generative study of second language acquisition*. Hillsdale, NJ: Erlbaum, 111–130.
Williams, J. N. & P. P. Lovatt (2003). Phonological memory and rule learning. *Language Learning* 53, 67–121.

ALAN JUFFS is Associate Professor in the Department of Linguistics, University of Pittsburgh. He is the Director of the English Language Institute at the University of Pittsburgh and co-editor of the Pitt Series in ESL textbooks published by the University of Michigan Press. He is currently President of the University and College Intensive English Programs Consortium (UCIEP). His research interests include the semantics–syntax interface and second language sentence processing. In addition to more theoretical aspects of SLA, he conducts classroom research in ESL vocabulary teaching and materials development.
MICHAEL HARRINGTON is a Senior Lecturer in Second Language Acquisition in the School of Languages and Comparative Cultural Studies, University of Queensland, Australia, where he teaches in the Master's Program in Applied Linguistics and TESOL Studies. His research interests are in second language processing with a particular interest in processing speed as a dimension of L2 proficiency development and performance. His current research focuses on the on-line assessment of vocabulary processing skill for placement and diagnostic purposes. He is also interested in the relationship between L2 proficiency and academic performance in the L2.

https://doi.org/10.1017/S0261444810000509 Published online by Cambridge University Press